THE PROTEIN SOCIETY

Amino-acid composition after loop deletion drives domain swapping

Neha Nandwani, Parag Surana, Jayant B. Udgaonkar,* Ranabir Das ,* and Shachi Gosavi*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India

Received 2 May 2017; Accepted 12 July 2017
DOI: 10.1002/pro.3237
Published online 14 July 2017 proteinscience.org

Abstract: Rational engineering of a protein to enable domain swapping requires an understanding of the sequence, structural and energetic factors that favor the domain-swapped oligomer over the monomer. While it is known that the deletion of loops between β-strands can promote domain swapping, the spliced sequence at the position of the loop deletion is thought to have a minimal role to play in such domain swapping. Here, two loop-deletion mutants of the non-domain-swapping protein monellin, frame-shifted by a single residue, were designed. Although the spliced sequence in the two mutants differed by only one residue at the site of the deletion, only one of them (YEIKG) promoted domain swapping. The mutant containing the spliced sequence YENKG was entirely monomeric. This new understanding that the domain swapping propensity after loop deletion may depend critically on the chemical composition of the shortened loop will facilitate the rational design of domain swapping.

Keywords: domain swapping; single-chain monellin; X-ray crystallography; hinge loop composition; hinge loop length

Abbreviations: MALS, multi-angle light scattering; MNEI, single chain monellin; SEC, size exclusion chromatography

Additional Supporting Information may be found in the online version of this article.

Neha Nandwani and Parag Surana contributed equally to this work.

Grant sponsor: Tata Institute of Fundamental Research and the Department of Science and Technology, Government of India, P.S. is supported by a Research Associate Fellowship from the Department of Biotechnology, India. J.B.U. is a recipient of a JC Bose National Fellowship from the Government of India. R.D. is a recipient of a DBT-Ramalingaswamy fellowship; Grant number: BT/HRD/23/02/2006; S.G. is supported by a grant from the Government of India-SERB; Grant number: EMR/2016/003885; and by a grant from the Simons Foundation to the Centre for the Study of Living Machines at NCBS, Bangalore.

**Correspondence to*: Shachi Gosavi, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, 560065, India. E-mail: shachi@ncbs.res.in; Ranabir Das, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, 560065, India. E-mail: rana@ncbs.res.in; and Jayant B. Udgaonkar, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, 560065, India. E-mail: jayant@ncbs.res.in

Introduction

Domain swapping is a mode of protein self-association between two or more polypeptide chains where intertwined oligomers are formed from monomeric proteins by the exchange of secondary or tertiary structural elements.[1–5] In this process, intramolecular interactions stabilizing the native conformation of a protein are replaced by nearly identical inter-molecular interactions. Consequently, the final structures of the folded monomer and of each of the units in the swapped multimer differ mostly with respect to the conformation of the hinge loop, the peptide segment through which the exchanging part is connected to the rest of the structure (Fig. 1). Depending on the swapped unit and mode of exchange, oligomers with versatile topologies can be generated *via* domain swapping.[3,6–11] Although engineered domain swapping has been used to regulate structural and functional properties in a few proteins,[7,12–21] its potential remains largely underutilized due to a lack of a clear understanding of the molecular mechanism and factors modulating domain swapping in proteins.[3,22]

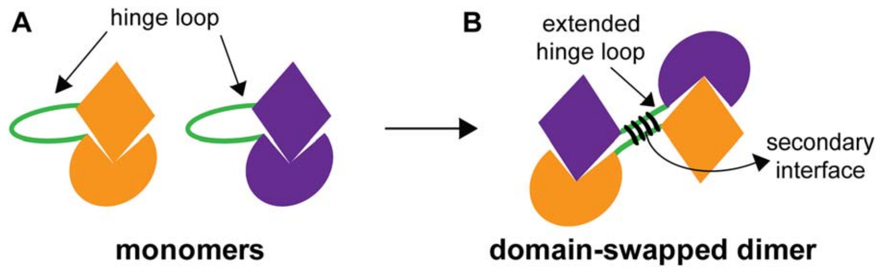

Figure 1. Schematic representation of domain swapping. **(A)** In the monomeric conformation, the hinge loop folds back upon itself, and folding proceeds intra-molecularly. **(B)** In the swapped conformation, the hinge loop adopts an extended conformation, and folding proceeds inter-molecularly. The overall fold of the protein is preserved. Each monomer-like structure formed by contributions from two polypeptide chains, is referred to as a functional unit. Sometimes, a secondary interface is created by the proximity of the two polypeptide chains in the swapped conformation.

Domain swapping has been introduced into several proteins that do not swap by modulating strain on the putative hinge loop.[4,23] The hinge loop converts from a loop or a turn in the monomer, to an extended conformation in the swapped multimers (Fig. 1). A conformationally strained hinge loop destabilizes monomeric structure, and results in swapped oligomers where the loop converts into a favorable extended conformation. Different ways of modulating strain on a protein include hinge loop shortening, lengthening, and mutations.[23] Loop shortening by 1–6 amino acid residues has been the most common approach to domain swapping design.[7,14,15,23–25] The importance of loop length is highlighted by the incremental increase in the proportion of swapped dimer observed for single chain Fv, caused by an incremental decrease in hinge loop length.[26] Shortening the hinge loop down to a size where it is geometrically difficult for the polypeptide to fold back on itself seems to be sufficient to induce domain swapping.

The amino acid composition of the spliced loop has received little attention while designing loop deletions, because it is believed not to play an active role in modulating the domain swapping propensity.[27] However, it has been shown that a single residue mutation in a tight turn can trigger domain swapping by changing the balance between the conformational strain in the hinge loop, and the entropic cost of dimerization.[25,28–30] This suggests that the spliced sequence at the site of deletion might contribute significantly to domain swapping propensity, because of possible synergism between strain due to the shorter loop length upon loop deletion, and the presence of amino acid residues that may increase backbone rigidity, adopt sub-optimal torsion angles, or may be unfavorable for turn formation.[31,32]

Here, we study the effect of the residue composition of the hinge loop, after loop deletion, upon the domain swapping propensity of the non-domain-swapping protein, single chain monellin (MNEI). MNEI is an α-β protein, with a β1-α1-loopA-β2-loop1-β3-loop2-β4-loop3-β5 topology [Fig. 2(A)]. MNEI is derived from the naturally occurring two-chain variant of monellin, dcMN, by the covalent linkage of the two chains (chain B: β1-α1-loopA-β2; chain A: β3-loop2-β4-loop3-β5) through a Gly-Phe linker.[33] MNEI is structurally homologous to the cystatin family of cysteine protease inhibitors.[34] Domain-swapped dimers have been observed for a few cystatins, in which almost half of the polypeptide chain is exchanged between the two molecules.[35–38] In these swapped dimers of the cystatins, loop1 opens up and forms a long, continuous β-strand, connecting the two "sub-domains" β1-α1-loopA-β2 (chain B of dcMN) and β3-loop2-β4-loop3-β5 (chain A of dcMN). The cystatin loop1 is on an average six residues shorter than loop1 in MNEI. Hence, we chose to study whether the shortening of loop1 would lead to the domain swapping of MNEI.

We created two different six-residue deletion variants of MNEI, frame-shifted by a single residue [Fig. 2(B)]. These variants, termed MNEI$\Delta 6^{\mathrm{Asn}}$ and MNEI$\Delta 6^{\mathrm{Ile}}$, differed by a single amino acid residue at the site of the deletion. We found that the

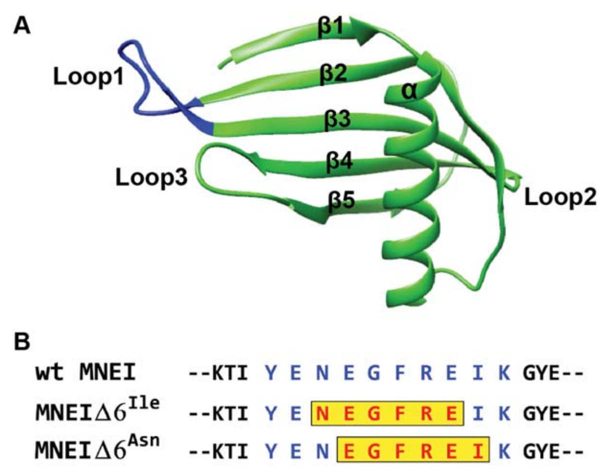

Figure 2. Design of MNEI loop deletion variants. **(A)** Structure of MNEI (PDB ID 1IV7) is shown in a cartoon representation. Different secondary structural elements and loops are labeled. Loop1 is highlighted in blue. **(B)** Residues 45–60 of MNEI are shown. Loop1 is composed of residues 48–57, which are indicated in blue. Residues deleted to generate each mutant variant are indicated in red.

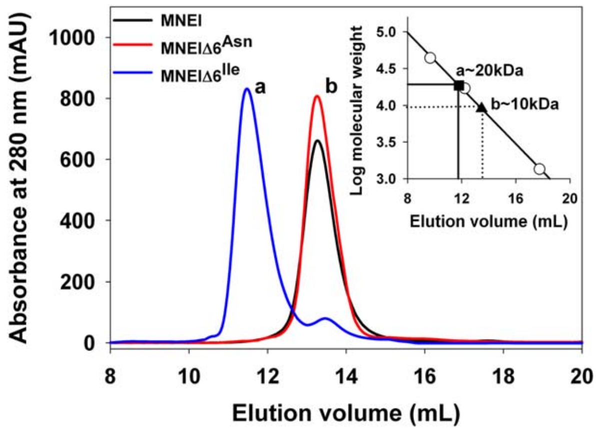

Figure 3. Frame-shifting by a single residue can lead to different protein fates. Size-exclusion chromatography profiles of MNEI and both the variants at pH 7, are shown. Elution volumes a and b correspond to those of the dimer and monomer, respectively, as calculated from the log molecular weight *versus* elution volume plot generated using molecular weight standards (inset).

oligomeric fates of the two variants were dramatically different. MNEIΔ6^{Asn} (containing the spliced sequence YENKG) is entirely monomeric in solution, similar to the wild type (wt) protein. The dominant species in MNEIΔ6^{Ile} (containing the spliced sequence YEIKG) is, however, a dimer, along with a barely detectable monomeric form. The structure of the MNEIΔ6^{Ile} dimer, determined to 2.6 Å resolution by x-ray crystallography, confirmed that dimerization was a result of domain swapping. In this structure, the two polypeptide chains that contribute to the MNEIΔ6^{Ile} dimer, exchanged sub-domains β1-α1-loopA-β2 and β3-loop2-β4-loop3-β5, similar to the cystatin swapped dimers. The striking effect of a single amino acid residue difference in the hinge region between MNEIΔ6^{Asn} and MNEIΔ6^{Ile}, on their oligomeric status indicates that the loop composition of the spliced region contributes significantly to the domain swapping propensity of a protein.

Results

Oligomeric fate of the loop-deletion variants of MNEI

Two different six-amino acid residue deletion mutant variants of MNEI were created such that they differed by a single amino acid residue at the site of deletion (see Materials and Methods). Figure 2(A) indicates the site of deletion mapped onto the tertiary structure of MNEI. Figure 2(B) highlights residues 50–55 and 51–56 that were deleted to generate the variants MNEIΔ6^{Ile} (spliced sequence YEIKG) and MNEIΔ6^{Asn} (spliced sequence YENKG), respectively.

Purified proteins were run on a size-exclusion chromatography column to analyze their oligomeric status in solution. MNEIΔ6^{Asn} was found to be entirely monomeric, similar to wt MNEI (Fig. 3). In contrast, MNEIΔ6^{Ile} was found to be predominantly dimeric, with the monomeric form corresponding to less than 10% of the population at a protein concentration as low as 5 μM. Multi-angle static light scattering experiments were carried out to determine the absolute molar mass of the dimeric fraction collected from the SEC column for MNEIΔ6^{Ile}. The apparent molecular weight (22,210 ± 460 Da) was in excellent agreement with that expected for a dimer (21,340 Da).

MNEIΔ6^{Ile} dimer is a domain-swapped dimer

MNEIΔ6^{Ile} dimer was crystallized, and its structure was solved by x-ray crystallography to a resolution of 2.6 Å (Supporting Information Table S1). It was evident that the MNEIΔ6^{Ile} dimer is formed by domain swapping [Fig. 4(A)]. The molecular replacement solution, obtained by using MNEI as a search model, revealed five polypeptide chains in the asymmetric unit of the crystal lattice, four of which dimerized among themselves while the fifth chain formed a swapped dimer with its symmetry related chain. A composite omit map calculated by simulated-annealing clearly showed continuous stretches of electron density, which fit well to a continuous β-strand stretch connecting the two units of the dimer [Fig. 4(B)], and supports a domain-swapped arrangement of the polypeptide chains. In contrast, modeling a turn into these calculated electron densities resulted in a poor fit and multiple steric clashes. The crystal structure revealed that the MNEIΔ6^{Ile} dimer is formed by the exchange of the β1-α1-β2 and β3-β4-β5 sub-domains, similar to the domain-swapped dimers reported for several cystatins[35–38] [Fig. 4(A)]. Other than loop1, which adopted an extended conformation in the swapped dimer [Fig. 4(C)], the overall fold and native contacts of monomeric MNEI were preserved in the MNEIΔ6^{Ile} swapped dimer [Fig. 5(A)]. Superposition of either of the halves of the swapped dimer (monomer subunits) with the MNEI monomer, excluding the hinge loop, yielded a root mean square deviation (rmsd) of 1.26 Å, indicating that the two conformations were nearly identical. The three swapped dimers observed in the crystal lattice differed only in a slight displacement about the symmetry axis in the dimer (Supporting Information Fig. S1). The dimers thus appeared to be flexible, which explains the high B-factors in the crystal structure (Supporting Information Table S2).

The secondary interface in domain swapping is a new intermolecular interface that is established due to the proximal arrangement of polypeptide chains in the swapped conformation (Fig. 1). This interface contributes to the stability of the dimer. A significant secondary interface was found to have been created in the MNEIΔ6^{Ile} swapped dimer. Loop1 opens up and forms a new β-strand that

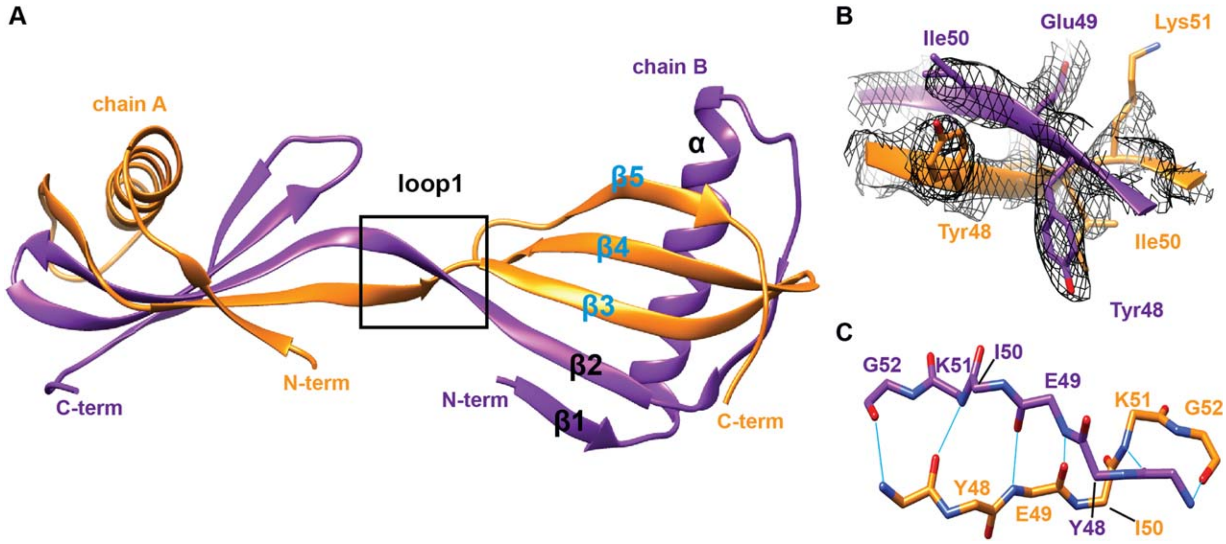

Figure 4. The MNEIΔ6$^{\mathrm{Ile}}$ dimer is a domain-swapped dimer. **(A)** Structure of the MNEIΔ6$^{\mathrm{Ile}}$ variant is shown. The two polypeptide chains contributing to the dimer are shown in purple and orange. The boxed region highlights loop1. Sub-domains, β1-α1-β2 and β3-β4-β5, contributed by the two polypeptide chains, are labeled in black and blue, for one of the functional units. **(B)** Zoomed in view of the region of cross-over between the two chains showing the YEIKG stretch in an extended conformation. The boxed region of (A) is shown with the simulated anneal composite omit electron density map at a contour level of 1.2 sigma. Residue side-chains are shown as sticks. Nitrogen atoms are colored blue and oxygen atoms are colored red. **(C)** The secondary interface composed of a β-sheet formed by the linker region from the two polypeptide chains is shown. Only the backbone traces of the two chains within the boxed region in (A) are shown. Backbone hydrogen bonds formed in the hinge region are shown by blue lines.

continues from the β2 strand into the β3 strand of each polypeptide chain in the dimer. As a result, a new anti-parallel β-sheet was formed by the two hinge loops [Fig. 4(C)] in the crossover region between the two polypeptide chains. This new interface created a long contiguous anti-parallel β-sheet, formed between the β2-loop1-β3 segments of the two chains that zipped the two polypeptides together *via* new hydrogen bonds [Fig. 4(C), Supporting Information Table S2]. Apart from the hydrogen bonds, multiple hydrophobic and vander Waals contacts stabilize the hinge region (Supporting Information Table S2). Such an arrangement limits the oligomerization *via* domain swapping to a dimeric species in MNEIΔ6$^{\mathrm{Ile}}$. Loop1 residues ^{48}YEIKG^{52} in the anti-parallel β-sheet conformation are arranged such that Ile50 of one chain stacks between the aromatic ring of Tyr48 of the other chain, and Tyr74

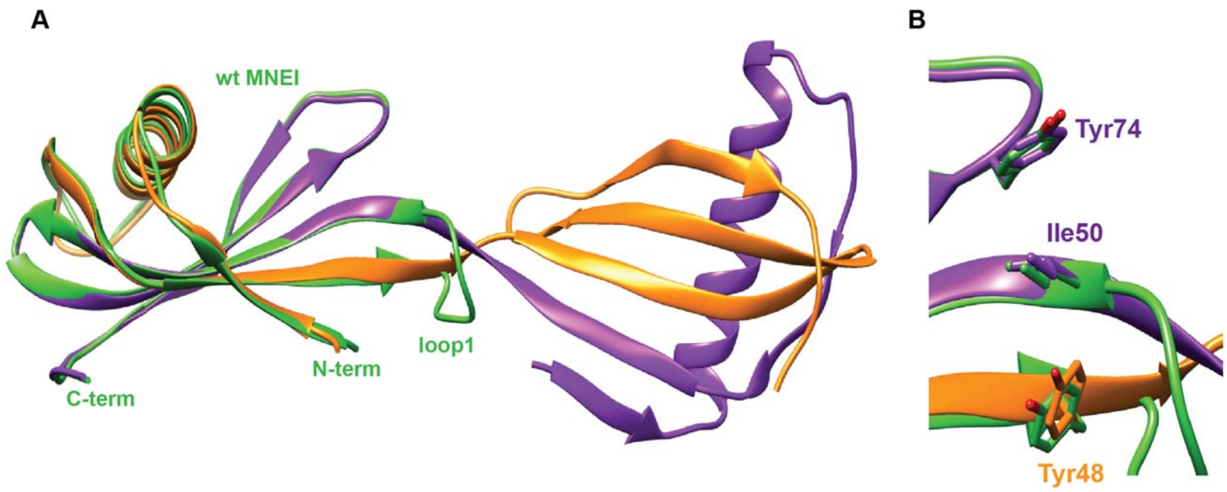

Figure 5. Similar stacking interactions in wt MNEI and MNEIΔ6$^{\mathrm{Ile}}$ swapped dimer. **(A)** The MNEI monomer (green) is aligned with the swapped dimer. For clarity, alignment with the other half of the dimer is not shown. **(B)** CH-π interactions between Y48, I56 and Y80 in the MNEI monomer occur inter-molecularly in the swapped dimer between Y48 from one chain and I50 and Y74 from another. The mismatch in residue numbers is due to renumbering after the site of deletion. For clarity, similar interactions in the other half of the swapped dimer are not shown.

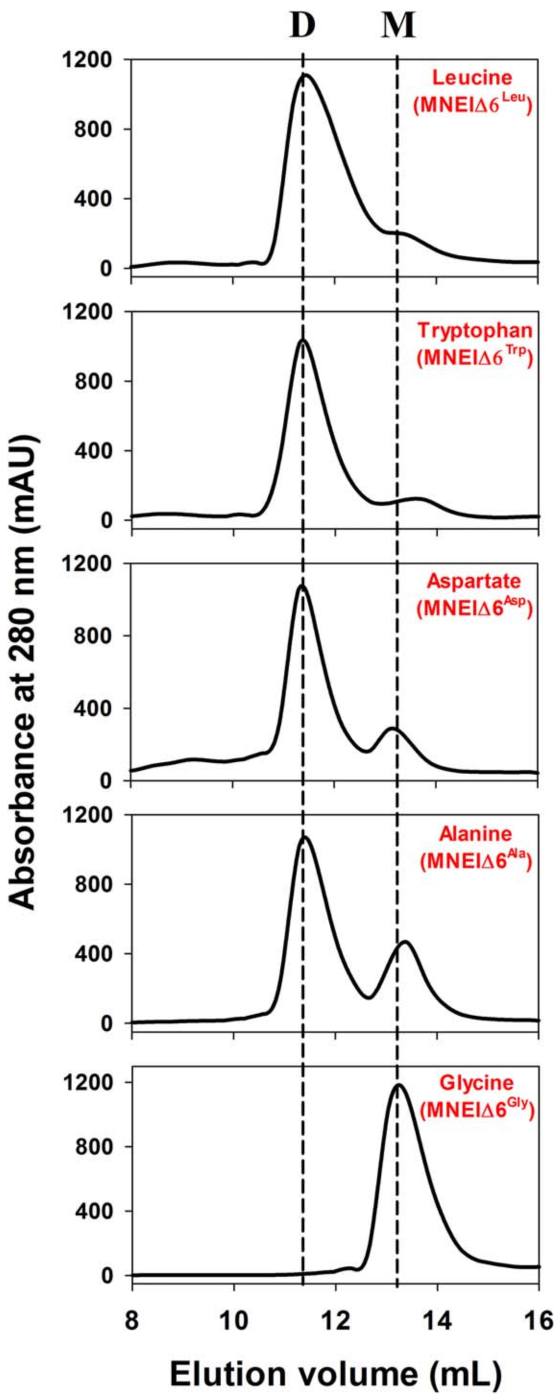

Figure 6. Domain swapping in other single site mutant variants of MNEIΔ6. Size-exclusion chromatography profiles of MNEIΔ6^{Leu}, MNEIΔ6^{Trp}, MNEIΔ6^{Asp}, MNEIΔ6^{Ala}, and MNEIΔ6^{Gly} at pH 7 are shown. The elution volumes of the dimer (D) and monomer (M) are indicated. The fraction of protein present in the dimeric form, calculated from the integrated area under the monomer (M) and dimer (D) peaks, is 86% (MNEIΔ6^{Leu}), 87% (MNEIΔ6^{Trp}), 80% (MNEIΔ6^{Asp}),70% (MNEIΔ6^{Ala}) and 0% (MNEIΔ6^{Gly}).

(from loop3) of the same chain [Fig. 5(B)]. In the wt MNEI monomer, a similar set of interactions exists between Tyr48, Ile56 (Ile50 in the dimer due to the loop deletion) and Tyr80 (Tyr74 in the dimer) [Fig. 5(B)]. Thus, domain swapping preserves the stacking interactions between these residues to bury the side chain of Ile50. A salt-bridge between Lys51 and Glu49 is observed in one dimer, but not in the other two.

Domain swapping in other single site mutant variants of MNEIΔ6

To better understand the difference in the domain swapping behavior of MNEIΔ6^{Ile} and MNEIΔ6^{Asn}, five other mutant variants of MNEIΔ6 were created, in which Ile/Asn was mutated to amino acid residues with different hydrogen-bonding capacities and/or hydrophobic nature. The oligomeric status of these mutant variants was determined. Mutation to Leu (MNEIΔ6^{Leu}) and Trp (MNEIΔ6^{Trp}), which are hydrophobic residues, promoted dimerization in MNEIΔ6 (Fig. 6). Similar to MNEIΔ6^{Ile}, both MNEIΔ6^{Leu} and MNEIΔ6^{Trp} were predominantly dimeric, with the monomeric form corresponding to less than 15% of the population at a protein concentration as low as 5 μM. Mutation to Asp (MNEIΔ6^{Asp}), which is a polar residue that is similar in size to Asn but has a distinct hydrogen-bonding pattern, also promoted dimerization (Fig. 6). Mutation to Ala (MNEIΔ6^{Ala}), which is neither β-branched (unlike Ile) nor capable of H-bonding (unlike Asn), was found to result in a mixture of the dimeric and monomeric forms (Fig. 6). Lastly, mutation to Gly (MNEIΔ6^{Gly}), which has the highest conformational flexibility and steric freedom, disfavored dimerization and MNEIΔ6^{Gly} was found to be entirely monomeric (Fig. 6).

Discussion

Mutations in MNEIΔ6, reduced loop strain and a reduced propensity to domain-swap

Hinge loops are likely to be strained in proteins that can domain-swap.[23,39] This conjecture is supported by the fact that the types of residues that occur in hinge loops in domain-swapping proteins are different from the types of residues that occur in the loops of non-domain-swapping proteins.[40] The intrinsic structural propensity of a residue determines whether it is locally stable in a given secondary structural element.[41] Deleting a loop which connects two secondary structural elements (here β-strands) is likely to create local strain in the folded structure by forcing residues that belong to the β-strands to form a loop or a turn. Strained loops can result in the loop residues occupying disallowed regions of the Ramachandran plot.[23,31,39] Studies of such disallowed regions have shown that certain types of residues (e.g. polar residues such as Asn) occur more often in such strained regions, and are thus likely to be able to tolerate such strain.[31] These residues tolerate strain either by forming side chain-main chain hydrogen bonds, or by creating local distortions in bonds and angles.[31,32]

However, hydrophobic residues such as Val and Ile rarely occupy disallowed regions of the Ramachandran plot and are expected to not be able to tolerate strain. The presence of hydrophobic residues in strained regions may lead to increased solvent-exposure of their side chains, because of the inability to adopt conformations favorable to the burial of these side chains.[42] In such cases, domain swapping may be favored not only because it relieves loop strain, but also because it aids the burial of the hydrophobic side chains.

It is known from the monomer structures of the cystatins that the ϕ and the ψ angles of a Val structurally homologous to the Ile in MNEI$\Delta6^{Ile}$ are such that the Val occurs in the disallowed region of the Ramachandran plot.[35,43–45] Further, a V$\rightarrow$N mutation in human cystatin C,[42] and a V$\rightarrow$D mutation in chicken cystatin[35] stabilizes the hinge loop, and reduces domain swapping. It is expected that a similar effect is at play in the two MNEI mutants, MNEI$\Delta6^{Ile}$ and MNEI$\Delta6^{Asn}$, and in fact the Ile of MNEI$\Delta6^{Ile}$ does get buried upon domain swapping [Fig. 5(B)]. In summary, the deletion of amino acid residues can cause local strain in the loops of proteins. The composition of the final spliced loop (after deletion) decides whether this local strain can be tolerated in a monomeric form of the protein. Bioinformatics studies suggest that polar amino acid residues are likely to be able to tolerate such local strain, while the presence of large hydrophobic amino acid residues is likely to drive domain swapping. This is potentially the main reason why MNEI$\Delta6^{Ile}$ domain-swaps, while MNEI$\Delta6^{Asn}$ does not. It can be hypothesized, on the basis of the MNEI$\Delta6^{Ile}$/MNEI$\Delta6^{Asn}$ dimerization propensities that placing a bulky, hydrophobic residue at the apex of a solvent-exposed, strained β-turn will result in domain swapping in proteins.

To further test this hypothesis, five single site mutations were made in loop1 in MNEI$\Delta6$ (Fig. 6). These mutations show that residues that are expected to increase local strain indeed promote dimerization in MNEI$\Delta6$. MNEI$\Delta6^{Gly}$, which has a glycine residue instead of Ile/Asn in loop1, is entirely monomeric. This is expected because Gly has a larger conformational space available to it, and is thus least likely to introduce local strain in the β-turn in the folded monomer. On the other hand, bulky hydrophobic residues, Leu and Trp promote dimerization in MNEI$\Delta6$. Leu is of the same size as Ile, but does not have a β-branched side chain. This suggests, in agreement with the cystatin results,[35,43–45] that the effect of Ile/Leu on the oligomeric status of MNEI$\Delta6$ might be solely due to the hydrophobicity of their side chains. It is expected that in addition to the hydrophobic nature of Trp, dimerization in MNEI$\Delta6^{Trp}$ might be promoted due to the preservation of stacking interactions in the dimer (Fig. 5). Similarly, dimerization is promoted in the Ala mutant (MNEI$\Delta6^{Ala}$), potentially because Ala is not tolerated in strained regions as well as a polar residue like Asn, in spite of the considerable steric freedom available to it.[31] Nevertheless, Ala is tolerated better than bulky hydrophobic residues, which is reflected in the increased relative population of the monomeric form in MNEI$\Delta6^{Ala}$. Surprisingly, mutation to Asp also promotes dimerization in MNEI$\Delta6$, unlike similar mutations in the cystatins.[35,46] Asp and Asn are polar residues which occur more often in strained regions;[31] however, these residues have contrasting effects on the oligomeric fate of MNEI$\Delta6$. Modeling Asp and Asn residues in the MNEI$\Delta6^{Ile}$ swapped dimer suggests that both Asp and Ile make similar number of intra- and interchain contacts, while Asn makes fewer contacts within the secondary interface. Moreover, Asn, but not Asp, can form an intra-chain hydrogen bond with the glycine residue at position 52, which can stabilize the monomeric conformation of the protein due to stabilization of the β-turn in the monomer. It is also possible that dimerization in MNEI$\Delta6^{Asp}$ is driven by charge-charge repulsion in the YEDKG stretch in loop1 in the monomeric form, which can be relieved upon dimerization and β-sheet formation in the hinge. It is important to note that such hypotheses about the mechanism and the driving force for dimerization in the mutant variants will need to be confirmed in future studies, by solving the structures of their monomeric and/or dimeric forms.

Loop composition and domain swapping

Both residue mutations and loop deletions can change the composition of the loop and increase loop strain to the point that domain swapping becomes energetically favorable.[3,4,23,27] As an example, a single amino acid mutation can drive domain swapping in the B1 domain of protein L.[28] The second β-hairpin turn is already strained in this protein, and the introduction of a nonglycine residue (G55A) within this strained conformation promotes domain swapping. Two additional mutations (A52V, N53P) also convert protein L to an obligate dimer.[47] In general, amino acid substitutions can increase conformational strain in loops by increasing steric clashes (large residues), exposing hydrophobic residues to the solvent, or restricting conformational space (prolines). In fact, prolines, being conformationally restricted, have been used to tune hinge loop strain and domain swapping in several proteins.[4,25,29,48–50] A closer look at proteins such as CD2 and suc1, where loop deletion has been shown to drive domain swapping, also reveals proline containing spliced sequences.[15,25] Two different non-overlapping three-residue deletion variants of the hinge loop in suc1 have been shown to dimerize to different extents, depending on whether or not prolines were retained in the hinge after deletion.[25] An example of loop deletion where domain swapping propensity cannot be attributed to prolines is staphylococcal nuclease.[14] Deletion of 6 residues in the putative hinge converted staphylococcal nuclease to a stable dimer in which the C-terminal helix swaps. The spliced sequence at the site of deletion was composed of residues that have a high normalized-propensity score of being in a hinge loop (deleted set: VYKPNN, spliced sequence: LAKVAYTH; Val, Ala, Tyr, Lys, Leu have high normalized-propensity scores, in that order).[40]

It is likely that the entropic cost of dimerization in domain swapping is not only countered by the release of conformational strain in the loop, but also by enthalpic stabilization of the dimer because of the formation of the secondary interface (Fig. 1). Hinge residue mutations in the W28A mutant of thioredoxin (Trx)[30] and the P43M mutant of calbindin D_{9k}[51] are hypothesized to promote domain swapping by stabilizing hydrophobic interactions at the secondary interface. In MNEI$\Delta6^{Ile}$ as well, the secondary interface comprising of several main chain hydrogen bonds likely further stabilizes the extended conformation of loop1 in the domain-swapped dimer.

The results presented here can be used to maximize the efficiency of engineering domain swapping in proteins by loop shortening. Shortening the putative hinge loop, such that the spliced sequence is composed of residues that rarely occupy disallowed regions of the Ramachandran plot, can increase the domain swapping propensity of a protein significantly. Further, the domain-swapped conformation is a compelling energy minimum on the folding landscape of a designed protein, because of the nearly identical sets of interactions that are at play in a monomer and a swapped dimer.[52] Tuning the hinge loop composition can be a simple strategy for the "negative design" of domain swapping in computationally designed proteins.

In summary, the present study reveals that although loop deletion introduces strain, and makes it difficult for the polypeptide to fold back onto itself, it is the composition of the spliced loop which finally determines domain swapping. Our data highlights the extent to which the monomer-dimer equilibrium can be impacted by the residue composition of the resultant hinge loop, suggesting that the spliced sequence at the site of deletion merits more attention while designing domain-swapped oligomers by loop shortening.

Materials and Methods

Construction of MNEI variants

The sequence of loop1 of MNEI is 48-YENEGFREIK-57, and Gly at position 58 immediately follows loop1. Deletion of the stretch NEGFRE (residues 50–55), which is centered on the Gly-Phe linker, results in the variant MNEI$\Delta6^{Ile}$, in which the spliced sequence is YEIKG. Similarly, deletion of the stretch EGFREI (residues 51–56) results in the variant MNEI$\Delta6^{Asn}$, in which the spliced sequence is YENKG. Incidentally, deletion of residues 49–54 (ENEGFR) recreated the variant MNEI$\Delta6^{Ile}$; hence, only two frame-shifted variants were constructed. Deletions were engineered by site-directed mutagenesis. Other single site variants of MNEI$\Delta6$ (MNEI$\Delta6^{Ala}$, MNEI$\Delta6^{Asp}$, MNEI$\Delta6^{Gly}$, MNEI$\Delta6^{Leu}$ and MNEI$\Delta6^{Trp}$) were generated by mutating Ile in MNEI$\Delta6^{Ile}$ to Ala, Asp, Gly, Leu, and Trp, respectively. Primers for site-directed mutagenesis were obtained from BioServe, India.

Assessment of the oligomeric status of MNEI variants

MNEI mutant variants were analyzed by analytical size-exclusion chromatography (SEC), using a Superdex 75 10/300 GL column on an ÄKTA FPLC, which resolves proteins in the molecular weight range 3–70 kDa. The column was run at 0.5 mL/min in 50 mM phosphate buffer, with 100 mM NaCl, 250 μM EDTA and 1 mM DTT, at pH 7. Protein elution was monitored at 280 nm. Molecular weights of the variants were estimated from a calibration curve generated using a Bio-Rad gel filtration standard.

Absolute molar mass and hydrodynamic radius determination for the MNEI variants was done using multi-angle light scattering on a DAWN 8+, eight angle light scattering instrument (Wyatt Technology Corp., Santa Barbara, CA). The concentration of the dimer isolated from the SEC was adjusted to approx. 1.5 mg/mL. Proteins were run through a 0.02 μm filter into the light scattering fused silica flow cell at a constant flow rate. A solution of monomeric bovine serum albumin was used for normalization of the scattering intensity. Data analysis was done using the software Astra.

Accession Numbers

Coordinates and the structure factor for MNEI$\Delta6^{Ile}$ have been deposited in the Protein Data Bank (PDB) under the accession codes 5XFU.

Author Contributions

N.N. and P.S. performed experiments. J.B.U. supervised the protein purification and characterization experiments, and R.D. supervised the crystallization experiments. All authors analyzed the data and wrote the manuscript.

Acknowledgments

We thank Nahren M. Mascarenhas for discussions, Praneeth Reddy for help with MNEI$\Delta6^{Asn}$ protein purification and Harish Kumar for help with MALS measurements. We thank S. Ramaswamy and Vinod Nayak for assistance with x-ray diffraction data

collection at the Soleil Synchrotron, France. The crystals were screened at the NCBS XRD Facility.

References

1. Bennett MJ, Choe S, Eisenberg D (1994) Domain swapping: entangling alliances between proteins. Proc Natl Acad Sci USA 91:3127–3131.
2. Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. Protein Sci 4:2455–2468.
3. Gronenborn AM (2009) Protein acrobatics in pairs – dimerization via domain swapping. Curr Opin Struct Biol 19:39–49.
4. Rousseau F, Schymkowitz J, Itzhaki LS (2012) Implications of 3D domain swapping for protein folding, misfolding and function. Adv Exp Med Biol 747:137–152.
5. Mascarenhas NM, Gosavi S (2017) Understanding protein domain-swapping using structure-based models of protein folding. Prog Biophys Mol Biol. 128:113–120.
6. Liu Y, Gotte G, Libonati M, Eisenberg D (2001) A domain-swapped RNase A dimer with implications for amyloid formation. Nat Struct Mol Biol 8:211–214.
7. Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF, Eisenberg D (2001) Design of three-dimensional domain-swapped dimers and fibrous oligomers. Proc Natl Acad Sci USA 98:1404–1409.
8. Liu Y, Gotte G, Libonati M, Eisenberg D (2002) Structures of the two 3D domain-swapped RNase A trimers. Protein Sci 11:371–380.
9. Gotte G, Helmy AM, Ercole C, Spadaccini R, Laurents DV, Donadelli M, Picone D (2012) Double domain swapping in bovine seminal RNase: formation of distinct N- and C-swapped tetramers and multimers with increasing biological activities. PLoS One 7:e46804.
10. Miyamoto T, Kuribayashi M, Nagao S, Shomura Y, Higuchi Y, Hirota S (2015) Domain-swapped cytochrome cb_{562} dimer and its nanocage encapsulating a $Zn–SO_4$ cluster in the internal cavity. Chem Sci 6: 7336–7342.
11. Sun Z, El Omari K, Sun X, Ilca SL, Kotecha A, Stuart DI, Poranen MM, Huiskonen JT (2017) Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. Nat Commun 8:14814.
12. Mossing MC, Sauer RT (1990) Stable, monomeric variants of (lambda) Cro obtained by insertion of a designed (beta)-hairpin sequence. Science 250:1712–1715.
13. Stott K, Blackburn JM, Butler PJ, Perutz M (1995) Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. Proc Natl Acad Sci USA 92:6509–6513.
14. Green SM, Gittis AG, Meeker AK, Lattman EE (1995) One-step evolution of a dimer from a monomeric protein. Nat Struct Mol Biol 2:746–751.
15. Murray AJ, Head JG, Barker JJ, Brady RL (1998) Engineering an intertwined form of CD2 for stability and assembly. Nat Struct Mol Biol 5:778–782.
16. Chen YW, Stott K, Perutz MF (1999) Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. Proc Natl Acad Sci USA 96:1257–1261.
17. López-Alonso JP, Diez-García F, Font J, Ribó M, Vilanova M, Scholtz JM, González C, Vottariello F, Gotte G, Libonati M, Laurents DV (2009) Carbodiimide EDC induces cross-links that stabilize RNase A C-dimer against dissociation: EDC adducts can affect protein net charge, conformation, and activity. Bioconjug Chem 20:1459–1473.
18. Reis JM, Burns DC, Woolley GA (2014) Optical control of protein–protein interactions via blue light-induced domain wwapping. Biochemistry 53:5008–5016.
19. Ha J-H, Karchin JM, Walker-Kopp N, Castañeda CA, Loh SN (2015) Engineered domain swapping as an on/off switch for protein function. Chem Biol 22:1384–1393.
20. Lin Y, Nagao S, Zhang M, Shomura Y, Higuchi Y, Hirota S (2015) Rational design of heterodimeric protein using domain swapping for myoglobin. Angew Chemie Int Ed 54:511–515.
21. Karchin JM, Ha J-H, Namitz KE, Cosgrove MS, Loh SN (2017) Small molecule-induced domain swapping as a mechanism for controlling protein function and assembly. Sci Rep 7:44388.
22. Nagarkar RP, Hule RA, Pochan DJ, Schneider JP (2010) Domain swapping in materials design. Pept Sci 94:141–155.
23. Rousseau F, Schymkowitz JWH, Itzhaki LS (2003) The unfolding story of three-dimensional domain swapping. Structure 11:243–251.
24. Ivanov D, Tsodikov OV, Kasanov J, Ellenberger T, Wagner G, Collins T (2007) Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. Proc Natl Acad Sci USA 104:4353–4358.
25. Rousseau F, Schymkowitz JWH, Wilkinson HR, Itzhaki LS (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc Natl Acad Sci USA 98:5596–5601.
26. Whitlow M, Filpula D, Rollence ML, Feng S-L, Wood JF (1994) Multivalent Fvs: characterization of single-chain Fv oligomers and preparation of a bispecific Fv. Protein Eng 7:1017–1026.
27. Schlunegger MP, Bennett MJ, Eisenberg D (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem 50:61–122.
28. O'Neill JW, Kim DE, Johnsen K, Baker D, Zhang KYJ (2001) Single-site mutations induce 3D domain swapping in the B1 domain of protein L from Peptostreptococcus magnus. Structure 9:1017–1027.
29. Miller KH, Karr JR, Marqusee S (2010) A hinge region cis-proline in ribonuclease A acts as a conformational gatekeeper for C-terminal domain swapping. J Mol Biol 400:567–578.
30. Garcia-Pino A, Martinez-Rodriguez S, Wahni K, Wyns L, Loris R, Messens J (2009) Coupling of domain swapping to kinetic stability in a thioredoxin mutant. J Mol Biol 385:1590–1599.
31. Gunasekaran K, Ramakrishnan C, Balaram P (1996) Disallowed Ramachandran conformations of amino acid residues in protein structures. J Mol Biol 264: 191–198.
32. Ramakrishnan C, Lakshmi B, Kurien A, Devipriya D, Srinivasan N (2007) Structural compromise of disallowed conformations in peptide and protein structures. Protein Pept Lett 14:672–682.
33. Tancredi T, Iijima H, Saviano G, Amodeo P, Temussi PA (1992) Structural determination of the active site of a sweet protein A 1H NMR investigation of pMNEI. FEBS Lett 310:27–30.
34. Murzin AG (1993) Sweet-tasting protein monellin is related to the cystatin family of thiol proteinase inhibitors. J Mol Biol 230:689–694.
35. Staniforth RA, Giannini S, Higgins LD, Conroy MJ, Hounslow AM, Jerala R, Craven CJ, Waltho JP (2001) Three-dimensional domain swapping in the folded and

molten-globule states of cystatins, an amyloid-forming structural superfamily. EMBO J 20:4774–4781.
36. Janowski R, Kozak M, Jankowska E, Grzonka Z, Grubb A, Abrahamson M, Jaskolski M (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. Nat Struct Mol Biol 8:316–320.
37. Kokalj SJ, Gunčar G, Štern I, Morgan G, Rabzelj S, Kenig M, Staniforth RA, Waltho JP, Žerovnik E, Turk D (2007) Essential role of proline isomerization in stefin B tetramer formation. J Mol Biol 366:1569–1579.
38. Valadares NF, Oliveira-Silva R, Cavini IA, Almeida Marques I, D'Muniz Pereira H, Soares-Costa A, Henrique-Silva F, Kalbitzer HR, Munte CE, Garratt RC (2013) X-ray crystallography and NMR studies of domain-swapped canecystatin-1. FEBS J 280:1028–1038.
39. Orlikowska M, Szymańska A, Borek D, Otwinowski Z, Skowron P, Jankowska E (2013) Structural characterization of V57D and V57P mutants of human cystatin C, an amyloidogenic protein. Acta Cryst D69:577–586.
40. Håkansson M, Svensson A, Fast J, Linse S (2001) An extended hydrophobic core induces EF-hand swapping. Protein Sci 10:927–933.
41. Bergdoll M, Remy M-H, Cagnon C, Masson J-M, Dumas P (1997) Proline-dependent oligomerization with arm exchange. Structure 5:391–401.
42. Mou Y, Huang P-S, Thomas LM, Mayo SL (2015) Using molecular dynamics simulations as an aid in the prediction of domain swapping of computationally designed protein variants. J Mol Biol 427:2697–2706.
43. Dehouck Y, Biot C, Gilis D, Kwasigroch JM, Rooman M (2003) Sequence-structure signals of 3D domain swapping in proteins. J Mol Biol 330:1215–1225.
44. Shingate P, Sowdhamini R (2012) Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. PLoS One 7:e39305.
45. Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–222.
46. Orlikowska M, Jankowska E, Kołodziejczyk R, Jaskólski M, Szymańska A (2011) Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C. J Struct Biol 173:406–413.
47. Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, Turk V (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J 9:1939–1947.
48. Martin JR, Craven JC, Jerala R, Kroon-Žitko L, Žerovnik E, Turk V, Waltho JP (1995) The three-dimensional solution structure of human stefin A. J Mol Biol 246:331–343.
49. Kolodziejczyk R, Michalska K, Hernandez-Santoyo A, Wahlbom M, Grubb A, Jaskolski M (2010) Crystal structure of human cystatin C stabilized against amyloid formation. FEBS J 277:1726–1737.
50. Kuhlman B, O'Neill JW, Kim DE, Zhang KYJ, Baker D (2001) Conversion of monomeric protein L to an obligate dimer by computational protein design. Proc Natl Acad Sci USA 98:10687–10691.
51. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O'Keefe BR, Boyd MR, Wlodawer A, Gronenborn AM (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. Structure 10:673–686.
52. López-Alonso JP, Bruix M, Font J, Ribó M, Vilanova M, Jiménez MA, Santoro J, González C, Laurents DV (2010) NMR spectroscopy reveals that RNase A is chiefly denatured in 40% acetic acid: implications for oligomer formation by 3D domain swapping. J Am Chem Soc 132:1621–1630.