doi:10.1016/j.jmb.2006.03.047 J. Mol. Biol. (2006) 359, 390–402

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structures of Ubiquitin Insertion Mutants Support Site-specific Reflex Response to Insertions Hypothesis

Debra M. Ferraro, Daniel J. Ferraro, S. Ramaswamy and Andrew D. Robertson*

Department of Biochemistry Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242 USA

*Corresponding author

We previously concluded that, judging from NMR chemical shifts, the effects of insertions into ubiquitin on its conformation appear to depend primarily on the site of insertion rather than the sequence of the insertion. To obtain a more complete and atomic-resolution understanding of how these insertions modulate the conformation of ubiquitin, we have solved the crystal structures of four insertional mutants of ubiquitin. Insertions between residues 9 and 10 of ubiquitin are minimally perturbing to the remainder of the protein, while larger alterations occur when the insertion is between residues 35 and 36. Further, the alterations in response to insertions are very similar for each mutant at a given site. Two insertions, one at each site, were designed from structurally homologous proteins. Interestingly, the secondary structure within these five to seven amino acid residue insertions is conserved in the new protein. Overall, the crystal structures support the previous conclusion that the conformational effects of these insertions are determined largely by the site of insertion and only secondarily by the sequence of the insert.

Keywords: indels; homology modeling; protein engineering; crystallography; ubiquitin

Introduction

Insertions and deletions, also called indels, are major contributors to evolutionary changes in proteins. Indel differences between human and chimpanzee genomes were recently found to total approximately 90 Mb, or 3% of the genomes. While point mutation differences between the genomes occurred with higher frequency, indels accounted for more than twice as many base-pair differences.[1] Insertions are responsible for about 8% of diallelic variations within the human genome.[2] A notable example of this is the expanded glutamine repeat that leads to Huntington's disease.[3]

Indels are useful in protein engineering and design. For example, new functions can be created with insertions. A novel sequence-specific nuclease activity was created[4] by replacing a loop of the engrailed homeodomain helix-turn-helix motif, which retains the DNA-binding properties of engrailed homeodomain, with the Ca^{2+}-binding loop of an EF-hand, which retains the Ca^{2+}-binding properties of the EF-hand motif.[5] Similar results were obtained by replacing the engrailed homeodomain helix-turn-helix loop with a Cu(II)-binding motif.[6] Indels are able to change protein structures in ways that point mutations alone cannot achieve, making them particularly useful for protein engineering.[7] However, the structural effects of indels are not understood well enough to be predicted, leading to difficulties in their use for rational protein design.

The challenges associated with understanding insertions and deletions are manifested in homology modeling: while insertions and deletions are known to occur most often in the regions between elements of secondary structure,[8] the structural consequences are difficult to predict.[9] Systematic investigation into how insertions change protein structure is likely to provide valuable information toward improving the capabilities of homology modeling with regard to insertions and deletions.

Previously, we created insertional mutants of ubiquitin using insertions from structurally homologous ubiquitin-like proteins of known

Present address: A. D. Robertson, Keystone Symposia, Drawer 1630, Silverthorne, CO 80498, USA.

Abbreviations used: NCS, non-crystallographic symmetry.

E-mail address of the corresponding author: andyr@keystonesymposia.org

structure.[10] The two sites chosen for insertion were between residues 9 and 10 (the 9-10 loop), and between residues 35 and 36 (the 35-36 loop) of ubiquitin. These variants were designed to test the hypotheses that insertions and deletions are major contributors to structural divergence within the ubiquitin-like protein family and that, consequently, insertions into ubiquitin would drive the structure towards that of the ubiquitin-like proteins from which the inserts were derived. As a test for sequence-specificity, other ubiquitin variants were created in which the inserts were derived from non-homologous positions in ubiquitin-like proteins. Additionally, an octaglycine insertion was introduced at each site as a control.

Changes in NMR chemical shifts were used as an indicator of changes in protein structure.[10] Judging by this criterion, insertions into the 9-10 loop produced minimal structural changes, and the modest changes were very similar for all five different insertion mutants. In contrast, insertions into the 35-36 loop showed a global effect on the conformation of the protein which, nonetheless, was similar for the five different insertion mutants at this position. We have termed this apparent site-dependent but sequence-independent effect a reflex response. Using previously established relationships between chemical shift and hydrogen bonding,[11,12] we formed hypotheses regarding the structural changes in the 35-36 loop mutants (Figure 1). We predicted that the helix bends away from the insertion site and that the hydrogen bonds connecting strands of the sheet lengthen near the insertion and are compressed away from the insertion.

To confirm the reflex response observed by NMR and to address the structural hypotheses for the 35-36 loop mutants, we have solved the crystal structures of four loop insertion mutants of ubiquitin: octaglycine insertion mutants in both loops; a homologous insertion mutant with residues QVRELVG from MoaD[13] inserted between residues 9 and 10; and a homologous insertion mutant with residues RWALA from MoaD inserted between residues 35 and 36. To investigate the ability of widely available structure prediction tools to capture the effects of insertions on protein structure, the crystal structures are compared to structures generated by some simple attempts at homology modeling on publicly available servers.

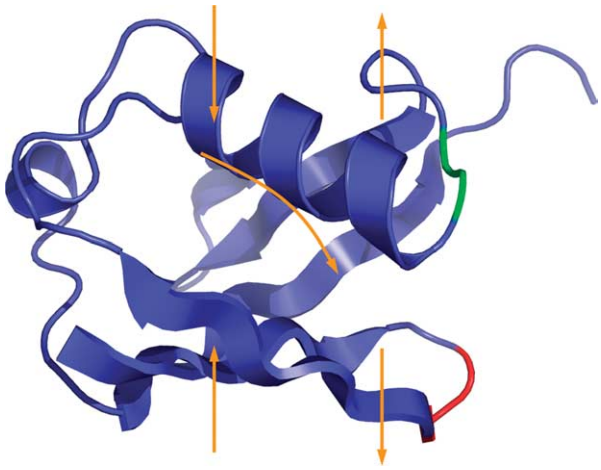

Figure 1. Predicted effects of insertions at the 35-36 loop on ubiquitin structure.[10] Residues 9 and 10 are shown in red and residues 35 and 36 are shown in green. The curved arrow indicates the predicted bend of the helix away from the 35-36 loop insertion site. The straight arrows indicate the predicted hydrogen bond lengthening of the sheet near the insertion site and shortening away from the insertion site.

Results

The nomenclature for the mutants discussed below indicates the location and composition of the insertion. For example, 9-10 G8 indicates that eight glycine residues were inserted between residues 9 and 10 of ubiquitin, and 35-36 MoaD was prepared by placing the structurally homologous insertion from MoaD between residues 35 and 36 of ubiquitin. The residue numbers discussed throughout Results and Discussion correspond to those of wild-type ubiquitin.

Selection and crystallization of mutants

In our previous study, insertions and their locations were chosen on the basis of structural homology.[10] All ten mutants studied by NMR[10] were screened for crystal formation. Four of the mutants crystallized: 9-10 G8, 9-10 MoaD, 35-36 G8 and 35-36 MoaD. Interestingly, both polyglycine mutants crystallized easily and none of the non-homologous mutants crystallized.

Determination of mutant structures

Crystals generally diffracted well; however, in some cases high lattice disorder resulted in very mosaic diffraction data. Because larger crystals tended to exacerbate this problem, small fragments of crystals were broken off from larger crystals and used for data collection. Due to the very small size of the crystals, synchrotron radiation was necessary to obtain high-resolution data.

For 9-10 G8, there were two molecules per asymmetric unit. The use of non-crystallographic symmetry (NCS) was not required in the refinement process because of the good quality of the data; consequently, the models for these two molecules were refined independently. In contrast, NCS restraints were required for refining the 9-10 MoaD model. The presence of thermal streaking in the diffraction data and high level of crystal mosaicity resulted in difficulty in integration of the spots. The low-resolution spots were the most affected and did not scale well. A low-resolution cut-off of 5 Å was used. Because of the use of NCS,