Journal of Molecular Graphics and Modelling 39 (2013) 118–125

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Parameterization of the proline analogue Aze (azetidine-2-carboxylic acid) for molecular dynamics simulations and evaluation of its effect on homo-pentapeptide conformations

Kyrylo Bessonov [1], Kenrick A. Vassall, George Harauz *

Department of Molecular and Cellular Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

ARTICLE INFO

Article history:
Received 18 July 2012
Received in revised form 11 October 2012
Accepted 17 November 2012
Available online 29 November 2012

Keywords:
Aze (azetidine-2-carboxylic acid)
Poly-proline II
cis–trans Isomerization
Molecular dynamics
GROMACS: myelin basic protein
Intrinsically disordered protein
Multiple sclerosis

ABSTRACT

We have parameterized and evaluated the proline homologue Aze (azetidine-2-carboxylic acid) for the gromos56a3 force-field for use in molecular dynamics simulations using GROMACS. Using bi-phasic cyclohexane/water simulation systems and homo-pentapeptides, we measured the Aze solute interaction potential energies, ability to hydrogen bond with water, and overall compaction, for comparison to Pro, Gly, and Lys. Compared to Pro, Aze has a slightly higher H-bonding potential, and stronger electrostatic but weaker non-electrostatic interactions with water. The 20-ns simulations revealed the preferential positioning of Aze and Pro at the interface of the water and cyclohexane layers, with Aze spending more time in the aqueous layer. We also demonstrated through simulations of the homo-pentapeptides that Aze has a greater propensity than Pro to undergo *trans*→*cis* peptide bond isomerization, which results in a severe 180° bend in the polypeptide chain. The results provide evidence for the hypothesis that the mis-incorporation of Aze within proline-rich regions of proteins could disrupt the formation of poly-proline type II structures and compromise events such as recognition and binding by SH3-domains.

1. Introduction

The non-protein imino acid azetidine-2-carboxylic acid (Aze, also denoted AZC in the literature) is similar to proline (2-carboxypyrrolidine, Pro), except that it has a 4-atom instead of a 5-atom ring (Fig. 1) [1,2]. This unusual imino acid was first discovered to occur naturally in plants of the *Liliaceae* family, specifically lily-of-the-valley (*Convallaria majalis*) [3]. It has since been found to occur naturally in many other plants, such as sugar beets (*Beta vulgaris*, formerly of *Chenopodiaceae*, now of *Amaranthaceae*) [4–7] and garden beets [8]. These "non-protein amino/imino acids" are important to plants in metabolic processes and as a defense mechanism [6,7,9,10].

When Aze is added endogenously to the media of cultured eukaryotic cells, it is recognized by the Pro-tRNA and is readily incorporated into proteins such as collagen, keratin, haemoglobin, and calnexin [8,11–18]. The homologue has also been shown to affect cell growth and proline metabolism when added to *E. coli*, both wild-type and proline auxotrophs [19–21]. We have recently shown that it could be mis-incorporated into *E. coli* expressing recombinant proteins such as myelin basic protein (MBP), but the viability of the cultures is decreased [22]. Some organisms such as yeast detoxify the compound by expressing *N*-acetyltransferases [23–25]. Other cells invoke a heat-shock response due to protein mis-folding [26–30].

The diverse toxic effects of Aze on animals are summarized by Rubenstein [31]. One example is the observation of severe neurodegeneration in sheep, including pregnant ewes that are fed a diet rich in beet by-products [32], and we note some other particular examples [33–35]. Rubenstein has further hypothesized that the misincorporation of Aze into proteins during brain development may contribute to inherently unstable myelin, which may be one of many factors predisposing an individual to developing multiple sclerosis (MS) [36,37]. Significantly, Aze enters the human food chain indirectly, since "beet molasses" from sugar production is used in many areas to feed cattle [17]. At present, the worldwide distribution of MS correlates with high rates of sugar beet production, *e.g.*, in the province of Alberta which has one of the highest rates of MS in Canada [38]. This hypothesis falls within the rubric of an "inside-out" model, which suggests that MS is caused by a cytodegenerative process aimed at the oligodendrocyte–myelin

Abbreviations: Aze, azetidine-2-carboxylic acid; MBP, myelin basic protein; GROMACS, GROningen MAchine for Chemical Simulations; MD, molecular dynamics; MS, multiple sclerosis; PDB, Protein Data Bank; PPI, poly-proline type I conformation; PPII, poly-proline type II conformation; SH3, Src homology 3.

* Corresponding author. Tel.: +1 519 824 4120x52535.
E-mail address: gharauz@uoguelph.ca (G. Harauz).
[1] Present address: Department of Electrical Engineering and Computer Science (Institut Montefiore), University of Liège, 4000 Liège 1, Belgium.

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.11.006

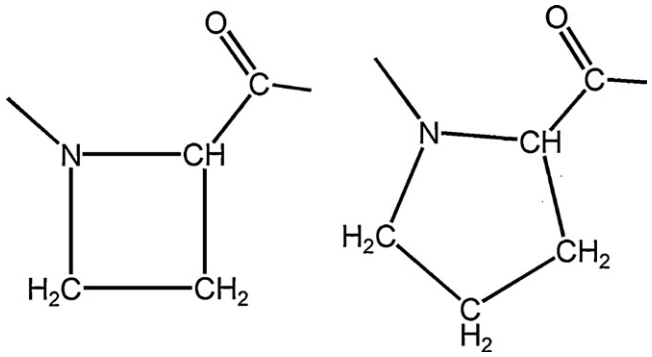

Fig. 1. Molecular structures of Aze (left) and Pro (right).

complex [39–41]. Gradual demyelination can then lead to an autoimmune response and a cycle of further degeneration, characteristic of the most common relapsing-remitting manifestation of MS.

Since the cause of MS remains unknown over 150 years since its first major clinical documentation [42], Rubenstein's hypothesis bears closer examination, including demonstrating the effects of Pro-to-Aze substitution in essential myelin proteins such as MBP [43–47]. This protein is multifunctional within myelin, acting both to adhere membrane leaflets together, and as a hub in protein–protein and protein–membrane interaction networks (actin and tubulin cytoskeleton, SH3-domain containing proteins, and calcium-activated calmodulin) [48–51]. We have determined that a central proline-rich region is functionally significant [46,50,52–56], and by molecular dynamics (MD) simulations have demonstrated that phosphorylation at key kinase targets affects its conformation, particularly a poly-proline (PPII) structure formed here [55]. We have recently conjectured that Pro-to-Aze mis-incorporation would have severe effects on the protein at the molecular level [22], particularly its interaction with SH3-domain-containing proteins [50–53,55,56]. Because of the difficulty, expense, and hazard of producing recombinant Aze-containing MBP for experimental investigations, it is worthwhile to pursue MD simulations. Towards this end, we present here a full parameterization of Aze for future MD studies using GROMACS [57,58], and illustrate its effects on the conformations of some simple homo-pentapeptides, in preparation for studies on larger protein segments. Since GROMOS force-fields are primarily used for protein simulations, accurate estimation of parameters for amino acids is crucial [59].

2. Methods

2.1. Parameterization of Aze topology in the GROMACS gromos53a6 force-field

Force-fields are essential components of any MD simulations, as they describe the simulation environment by assigning physical constants and empirical charges to all chemical species contained in a particular system. In the case of the GROMOS force-fields, most of the parameters are empirically determined, which makes parameterization of new chemical species, particularly the accurate estimation of atomic partial charges, a laborious task. Here, we decided to use a semi-automatic approach to parameterize Aze accurately. The structure of Aze was derived from that of Pro by elimination of the CH_2 group from the aliphatic ring, which facilitated the validation process (Fig. 1). The PRODRG2 online server [60] in conjunction with the Pro reference parameters was used to derive the Aze parameters for bonds, dihedral angles, improper dihedral and other angles, as discussed in detail in the Supplementary Information.

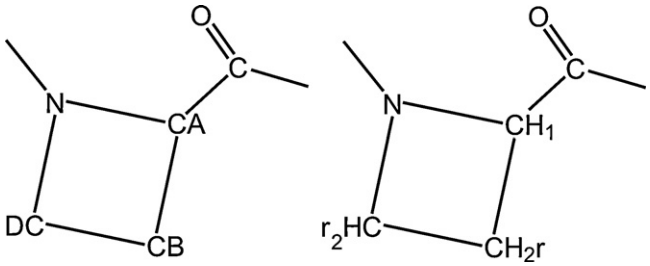

Fig. 2. Representations of Aze using the default Protein Data Bank (left) and gromos56a3 (right) nomenclatures. Please refer to Fig. S1 for the detailed Aze topology interpretation in the context of the gromos56a3 force-field.

Improper and proper dihedral angles of the Aze topology were almost identical to those of Pro, due to structural similarity. The only small modification was defined by the <(CA CB CD N) and <(N CA CB CD) angles, part of the 4-membered ring defined under gromos53a6 dihedral torsion type angles 34 (gd_34) (Fig. 2 and Fig. S1).

2.2. Validation of Aze parameters by QM calculations

The partial charges predicted by the PRODRG server were checked by *ab initio* QM calculations since accurate derivation of partial charges is essential, as discussed in Ref. [61]. The Hartree–Fock *ab initio* model with the 6-311G basis set of wave functions describing the atomic orbitals was used in all QM calculations. The final calculation of partial charges also employed an electro-static potential (ESP) fitting that considers interaction energies [62]. The overall charge of zero and electron spin multiplicity of one were specified for both the Aze and Pro residues. The QM calculations were performed on a single residue with uncharged termini, to represent more closely the context of the residue being part of a polypeptide chain (Fig. S2). The QM runs were done both for Pro and Aze, for comparison purposes. The charges assigned by Gromos53a6 were taken as a reference. The Gaussian 09 software package was used for all QM calculations [http://www.gaussian.com/g_tech/g_ur/m_citation.htm; Gaussian, Inc., Wallingford, CT].

2.3. Molecular dynamics simulations of Aze-containing peptides

The evaluation of the behaviour of Aze within GROMACS was performed by molecular dynamics simulations of homo-pentapeptides in aqueous and mixed solvent (cyclohexane/water) systems. Molecular dynamics simulations were performed using the GROMACS 4.5.5 software package [57,58] and the Gromos96 ffG53a6 force-field [59], at 310 K and 1 bar for total of 20 ns. Molecular dynamics production runs were performed using 64 processors at the Compute Canada/SharcNet facilities (https://www.sharcnet.ca). For visualization and analysis of structural files and trajectories computed by GROMACS, the GROMACS utilities and the Visual Molecular Dynamics (VMD) program [63] were utilized. The custom scripts for VMD were written in the TCL language to compute the φ, ψ, and ω dihedral angles. The frequency of conversion of the peptide bond from *cis* to *trans* conformation as well as the duration of time spent in each conformation was evaluated using custom Perl scripts.

2.4. Evaluation of the Aze topology parameters using bi-phasic and pure aqueous systems

The derived Aze topology parameters were evaluated in bi-phasic interfacial systems consisting of non-polar cyclohexane and polar spc216 water solvents. Such systems allow visualization of