ARTICLE

Structure of hen egg-white lysozyme solvated in TFE/water: a molecular dynamics simulation study based on NMR data

Andreas P. Eichenberger · Wilfred F. van Gunsteren · Lorna J. Smith

Abstract Various experimental studies of hen egg white lysozyme (HEWL) in water and TFE/water clearly indicate structural differences between the native state and TFE state of HEWL, e.g. the helical content of the protein in the TFE state is much higher than in the native state. However, the available detailed NMR studies were not sufficient to determine fully a structure of HEWL in the TFE state. Different molecular dynamics (MD) simulations, i.e. at room temperature, at increased temperature and using proton–proton distance restraints derived from NMR NOE data, have been used to generate configurational ensembles corresponding to the TFE state of HEWL. The configurational ensemble obtained at room temperature using atom-atom distance restraints measured for HEWL in TFE/water solution satisfies the experimental data and has the lowest protein energy. In this ensemble residues 50–58, which are part of the β-sheet in native HEWL, adopt fluctuating α-helical secondary structure.

Keywords Hen egg white lysozyme · TFE/water solvent · Protein structure · NMR NOE proton–proton distances · Molecular dynamics simulation

A. P. Eichenberger · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, Zurich, Switzerland
e-mail: wfvgn@igc.phys.chem.ethz.ch

L. J. Smith
Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, UK

Introduction

2,2,2-Trifluoroethanol (TFE) and related co-solvents have been widely used in the study of protein structure, folding and misfolding (Buck 1998). TFE is particularly recognised for its ability to stabilise α-helical secondary structure (Nelson and Kallenbach 1986; Jasanoff and Fersht 1994; Shiraki et al. 1995; Cammers Goodwin et al. 1996). However, its action depends on the peptide or protein sequence, the concentration of TFE and on the solution conditions used. Under certain conditions, TFE has also been seen to stabilise β-hairpins, turns and hydrophobic clusters, to accelerate protein folding, to facilitate the incorporation of peptides in membranes and to promote the formation of amyloid fibril aggregates (Buck 1998; Blanco et al. 1994; Smith et al. 1994; Lu et al. 1997; Chiti et al. 2000; Fezoui and Teplow 2002). Industrially, TFE is employed as solvent in catalysis (Bégué et al. 2004; Shuklov et al. 2007).

TFE is thought to enhance intra-solute hydrogen bonding particularly in secondary structure regions, whilst weakening long-range tertiary structure of the protein (Buck 1998; Cammers Goodwin et al. 1996; Rajan and Balaram 1996). Solution X-ray scattering studies suggest that an important factor is the clustering of the alcohol molecules in TFE aqueous solution (Hong et al. 1999). NMR data, supported by MD simulations of peptides in explicit TFE/water solutions, show that this clustering results in an accumulation of TFE molecules around the peptide surface (Fioroni et al. 2002; Roccatano et al. 2002; Diaz et al. 2002; Mehrnejad et al. 2007). By coating the peptide surface they partially exclude water molecules (Starzyk et al. 2005). It is suggested that this provides a local hydrophobic environment that promotes intrapeptide hydrogen bonding and maintains local hydrophobic interactions. For proteins an

Table 1 Overview the six MD simulations

Simulation name	pH	Solvent numbers and type	Initial configuration	T (K)	NOE distance restraints	Length (ns)
MD_pH7_H2O	7	14′355 H_2O 8 Cl^-	X-ray_EM[b]	310	–	20
MD_pH2_H2O	2	19′518 H_2O 19 Cl^-	X-ray_EM	310	–	20
MD_pH2_TFE	2	5′411 H_2O 19 Cl^- 3′138 TFE[a]	X-ray_EM	310	–	20
MD_pH2_TFE_HT	2	5′411 H_2O 3′138 TFE 19 Cl^-	MD_pH2_TFE[c]	310/400/310[d]	–	11.4/6/2.6[d]
MD_pH2_TFE_DR	2	5′411 H_2O 3′138 TFE 19 Cl^-	MD_pH2_TFE	310	NOE_TFE[e]	20
MD_pH2_TFE_HT_DR	2	5′411 H_2O 3′138 TFE 19 Cl^-	MD_pH2_TFE_HT_400[f]	400/310[g]	NOE_TFE	2/18[g]

[a] 70 % TFE/30 %H_2O (v/v)

[b] Energy-minimised X-ray structure 1AKI (Artymiuk et al. 1982)

[c] Final configuration of simulation "MD_pH2_TFE"

[d] T is increased from 310 to 400 K over 11.4 ns, then decreased to 310 K over the next 6.0 ns and then kept at 310 K for 2.6 ns

[e] See Table 2

[f] Configuration after 11.4 ns of simulation MD_pH2_TFE_HT

[g] T is decreased from 400 to 310 K over 2 ns and then kept at 310 K for 18 ns

Fig. 1 The different, sequential HEWL simulations. All initial coordinates have been generated from one of the two PDB files with PDB entry code 1AKI (Artymiuk et al. 1982) or 1E8L (Schwalbe et al. 2001) (*black ellipses*). For each simulated system (*rectangles*) the type of solvent and pH/protonation state/temperature/application of NOE distance restraints is indicated. Diamonds indicate the length of the simulation as well as the simulation temperature: *blue* 310 K; *blue →red* temperature increase from 310 to 400 K; *red →blue* cooling from 400 to 310 K. Consecutive simulations can be identified following the flow chart. See also Table 1 for more details
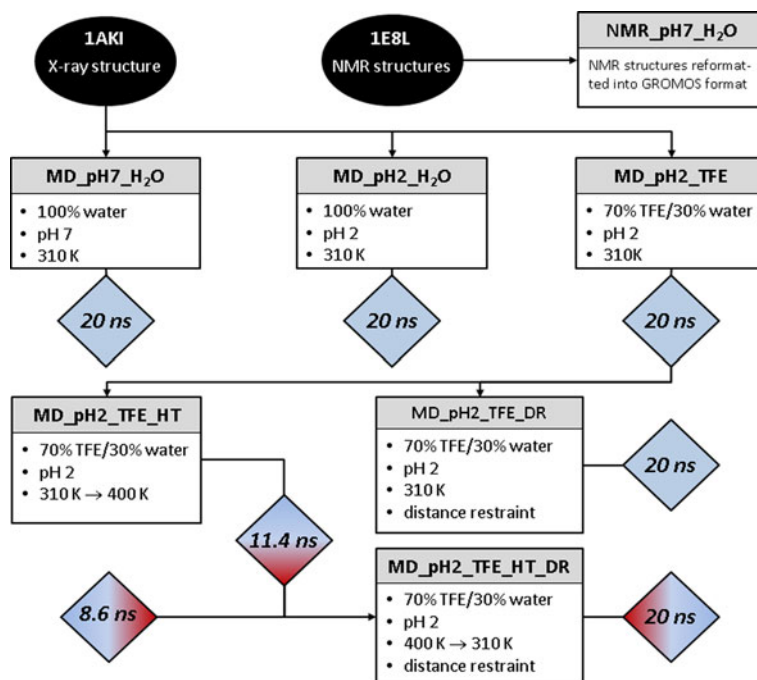
additional significant effect appears to be the penetration of TFE molecules into the protein interior (Kumar et al. 2003).

Hen egg white lysozyme (HEWL) is one of the proteins whose conformation has been studied extensively in TFE solutions (Buck et al. 1993, 1995, 1996; Yang et al. 1995; Hoshino et al. 1997; Povey et al. 2007). Low concentrations of added TFE co-solvent (<10 % (v/v)) stabilise the native structure of the protein (Buck et al. 1993;

Povey et al. 2007). However, further addition of TFE induces a cooperative transition to an intermediate state, referred to as the TFE state. The TFE state of HEWL contains substantial helical secondary in the absence of extensive persistent tertiary structure (Buck et al. 1993, 1995; Povey et al. 2007; D'Amico et al. 2011). Studies by SAXS show that the protein has an increased radius of gyration relative to the native state, although it is still relatively compact (Hoshino et al. 1997). A particularly interesting feature is that the helical content of the protein in the TFE state is higher than that in the native state (Buck et al. 1993; Yang et al. 1995). The native state structure of HEWL consists of an α-domain which contains four main α-helices (A-D) and a C-terminal 3_{10}-helix and a β-domain which contains a triple-stranded antiparallel β-sheet and a long loop. Detailed NMR studies show that in 70 % TFE/30 % water the regions corresponding to native helices in HEWL retain persistent helical conformation and increase in length, being extended at their termini. In addition, the C-terminal region of the β-sheet undergoes a conformational transition and adopts a more flexible helical structure for residues 50–58 (Buck et al. 1995; Yang et al. 1995).

Despite detailed study, the NMR data available for the TFE state of HEWL are not sufficient on their own to define a structure for this state. In any case the data suggest that the TFE state will consist of an ensemble of interconverting partially folded conformers rather than a single well-defined structure (Buck et al. 1995, 1996). In this work we have therefore used an MD simulation approach to attempt to generate an ensemble of structures to describe the TFE state of HEWL. A set of MD simulations of HEWL in an explicit solvent of 70 % TFE/30 % water molecules have been run, including simulations at elevated temperature and simulations with experimentally derived NOE distance restraints (Buck et al. 1995; Buck 1994). Taken together these simulation trajectories provide a

description of this conformationally diverse, partially folded TFE state. It is not the purpose of the present study to come up with a general mechanism for TFE induced protein conformational changes, which would require the simulation of a variety of proteins in a series of TFE/water mixtures of varying composition, a too expensive endeavour.

Materials and methods

Simulation software and force-field parameters

All molecular dynamics (MD) simulations and the analysis of the simulation trajectories described in this work were performed using the GROMOS software package for (bio)molecular simulations (Schmid et al. 2011, 2012; http://www.gromos.net) in combination with the GROMOS 53A6 force field (Oostenbrink et al. 2004). Water was modelled using the simple-point-charge (SPC) model (Berendsen et al. 1981) and TFE using the model of Fioroni et al. (Fioroni et al. 2000). Six configurational ensembles of the protein were generated by MD simulations under different conditions, see Table 1 and Fig. 1. The set of 50 NMR model structures that had been derived from NMR data of HEWL in aqueous solution (Schwalbe et al. 2001) and are available in the PDB with entry code 1E8L were reformatted into GROMOS format and will be indicated as NMR_pH7_H_2O.

Simulation protocol

Initial coordinates of the the six different lysozyme simulations (see Fig. 1; Table 1) were generated from the X-ray structure deposited at the Protein Data Bank (PDB) with PDB entry code 1AKI (Artymiuk et al. 1982). Missing

Table 2 Six NMR NOE proton–proton distance bound sets used as distance restraints in some simulations or for analysis of the set of NMR structures and MD generated configurational ensembles

Name	Number of NOE distance bounds				Description
	Total	Short	Medium	Long	
NOE_H_2O	1,630	999	278	353	NOE distance bounds of native HEWL in water (Schwalbe et al. 2001)
NOE_H_2O_NTFE	1,209	622	234	353	a subset of NOE_H_2O with NOE distance bounds not present in NOE_TFE
NOE_H_2O_NTFE_42-60	134	85	18	31	a subset of NOE_H_2O_NTFE with NOE distance bounds involving at least one of the atoms within residues 42–60
NOE_TFE	1,388	884	492	12	NOE distance bounds of HEWL in TFE/water (Buck 1994)
NOE_TFE_NH_2O	960	501	447	12	a subset of NOE_TFE with NOE distance bounds not present in NOE_H_2O
NOE_TFE_NH_2O_42-60	128	78	50	0	a subset of NOE_TFE_NH_2O with NOE distance bounds involving at least one of the atoms within residues 42–60

The first three sets are specified in Table S1 and the second three in Table S2. Pairs (i, j) are separated into short $(i \leq j \leq i + 1)$, medium $(i + 2 \leq j \leq i + 4)$, and long $(j \geq i + 5)$ ranged along the residue sequence indicated by i and j
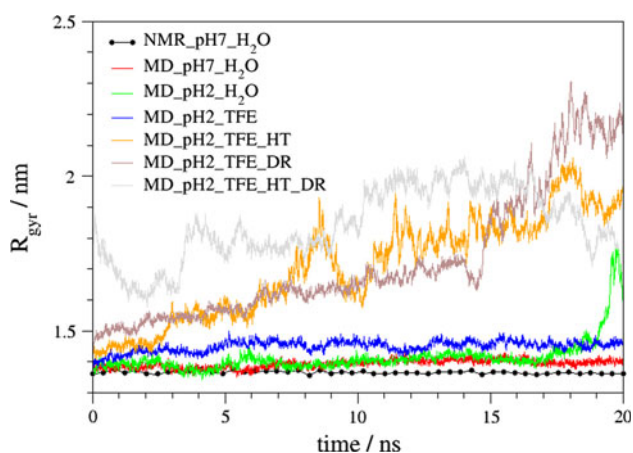
Fig. 2 Radius of gyration calculated for the backbone atoms of the seven sets of protein configurations. The simulation names are defined in Table 1 and Fig. 1. Note that for NMR_pH7_H$_2$O the time scale does not have any meaning but the 50 configurational NMR model structures were evenly distributed over the 20 ns time scale length of the X-axis

hydrogen positions were generated based on geometric criteria using the GROMOS++ program gch (Eichenberger et al. 2011). The protonation states of the amino acid residues were chosen corresponding to pH 2 or pH 7 (Table 1; Fig. 1). Histidine 15 was protonated according to its hydrogen-bonding environment and the pH value of the environment: at N$_\varepsilon$ for pH 7 and doubly protonated for pH 2 and all other HEWL structures. The four disulfide bridges were treated as covalently linked Cys residues. The different HEWL structures were energy minimised in vacuo and solvated in cubic boxes containing only water (Berendsen et al. 1981) molecules or a mixture of 30 % water/70 % TFE molecules, such that the minimum distance of a non-hydrogen protein atom to the box wall was 1.4 nm. This led to a box-edge length of 8.5 nm. At pH 7 HEWL has a net charge of +8 e, while at pH 2 this is +19 e. To yield an overall charge neutrality of the periodic boxes, 8 water molecules were replaced by chloride ions in the simulations at pH 7 and 19 water molecules at pH 2. The replacement was done using the GROMOS++ program ion (Eichenberger et al. 2011). The solvent was energy minimised with the positions of the heavy atoms within the protein kept fixed. Initial velocities for the three simulations starting from the energy-minimised X-ray structure were assigned from a Maxwell-Boltzmann distribution at 60 K with position restraining of the protein atoms with an initial harmonic force constant of 25,000 kJ mol^{-1} nm^{-2}. The temperature was stepwise raised while the force constant of the position restraining was lowered, ending in a non-restrained protein simulation at 310 K after 100 ps of simulation time. A 20 ns MD simulation followed this 100 ps thermalisation procedure. The protein configurations were saved for analysis every 0.5 ps.

All but the HT simulations were held at constant temperature and pressure (310 K, 1 atm) using the weak coupling algorithm (Berendsen et al. 1984) with corresponding coupling times of $\tau_T = 0.1$ ps and $\tau_p = 0.5$ ps, respectively. The protein and solvent were separately coupled to the heat bath. An estimated isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$ and 2.227×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$ was used for systems with the pure water solvent and TFE/water mixture, respectively. All bond lengths and the bond angles of the water molecules were kept rigid by applying constraints using the SHAKE algorithm (Ryckaert et al. 1977) with a relative geometric tolerance of 10^{-4}, allowing for an integration time step of 2 fs when solving the equations of motion using the leap-frog algorithm (Hockney and Eastwood 1981). Nonbonded (van der Waals and electrostatic) interactions were handled adopting triple-range cut-off radii: interactions within the short-range cutoff of 0.8 nm were calculated every time step from a pair list that was generated every five steps, when also interactions between 0.8 and 1.4 nm were computed. The long-range electrostatic forces were represented by a reaction field with a relative permittivity (Heinz et al. 2001) of $\varepsilon_{RF} = 61$ outside the long-range cutoff of 1.4 nm for the pure water solvent simulations. The simulations in TFE/water used a smaller relative permittivity of $\varepsilon_{RF} = 31$. The centre of mass translation and rotation were removed every 2 ps to avoid a flying ice cube (Harvey et al. 1998).

The three simulations MD_pH2_TFE_HT, MD_pH2_TFE_DR and MD_pH2_TFE_HT_DR used the same simulation parameters as described above but the initial atom positions and atom velocities were read from the final configuration of a preceding simulation. The high temperature simulations differ from the description above by a linear heating and/or cooling between 310 and 400 K. In the DR simulations distance restraints with a harmonic force constant of 2,500 kJ mol^{-1} nm^{-2} were applied to atom pairs from a set of NOE distance bounds derived from experiments. The six different sets of NMR NOE proton–proton distance bounds are specified in Table 2 and described below.

Sets of derived NMR NOE proton–proton distances

Two sets of NOE distance upper bounds derived from NMR experiments of HEWL in aqueous solution (Schwalbe et al. 2001) and the other from NMR experiments on HEWL in TFE/water solution (Buck 1994) were converted to GROMOS format and pseudo-atom bound corrections were added, see Table 2. A detailed description of the computational procedures with equations and distance corrections can be found in (van Gunsteren et al. 1996). 1630

experimental NMR proton–proton upper distance bounds were taken from (Schwalbe et al. 2001) and 1388 from (Buck 1994) including pseudo-atom distance corrections as given by Wüthrich et al. (1983). Proton–proton distances were calculated using $1/r^3$ averaging, $\overline{r} = (\langle r^{-3} \rangle)^{-1/3}$. The set NOE_TFE was used as distance restraints in simulations DR, Table 1. Different subsets of the two basic sets of NOE bounds were used for particular analyses.

Analysis

All analyses of the sets of protein structures have been done using the tools of GROMOS++ (Eichenberger et al. 2011). In particular, the following GROMOS++ programs have been used: rmsd (atom-positional root-mean-square deviation), rmsf (atom-positional root-mean-square fluctuations), rgyr (radius of gyration), dssp (detection of secondary structure in the protein), prep_noe, noe and post_noe (atom-atom distance bound violations).

When comparing or averaging quantities Q that depend on the position of the centre of mass and the spatial orientation of a particular set of atoms, the centres of mass are
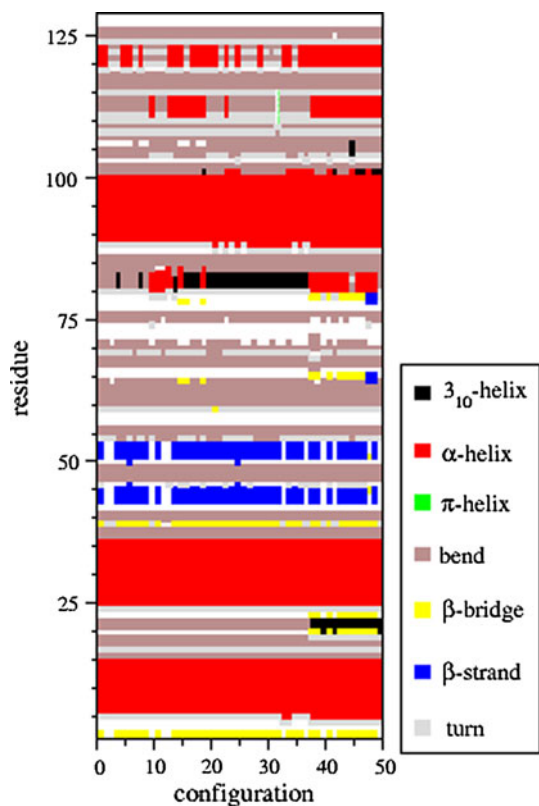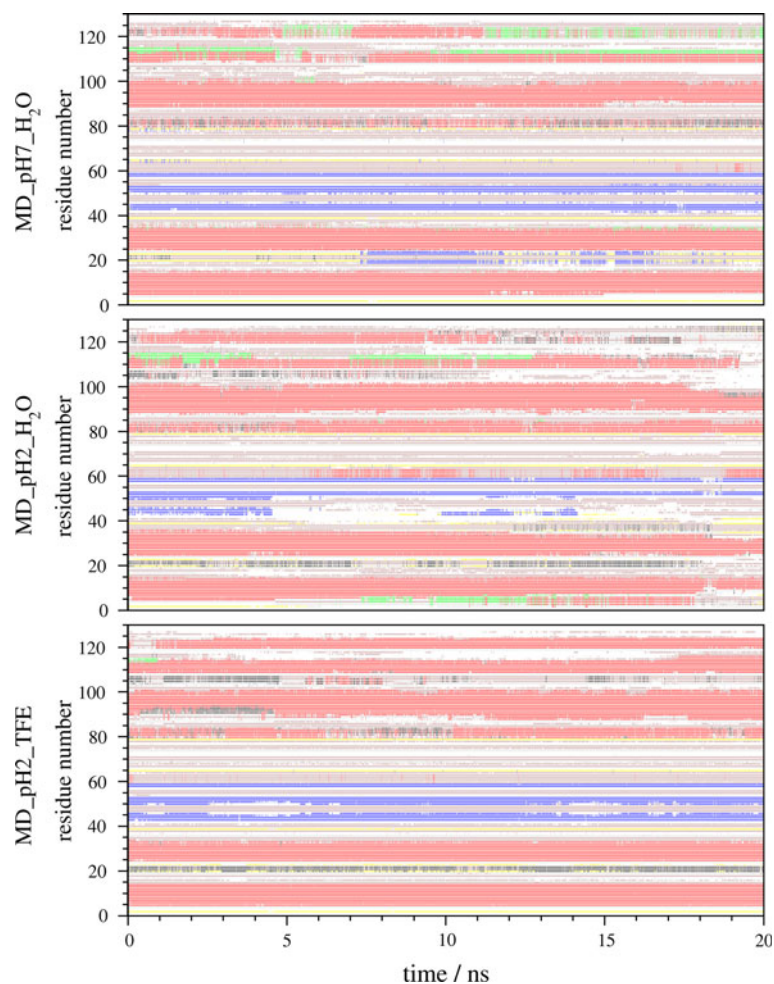

Fig. 3 Secondary structure elements (Kabsch and Sander 1983) calculated for the 50 configurations of the set of NMR model structures (Schwalbe et al. 2001) derived from NMR experiments of HEWL in aqueous solution at pH 7

superimposed and a rotational least-squares fit of the positions of the set of atoms is performed before Q is calculated. The rules of Kabsch and Sander (1983) have been applied to detect and monitor secondary structure elements in the HEWL simulations. In some cases one residue may be assigned to be part of two different secondary structure elements. In order to avoid ambiguous assignments in such cases, the following priority rules were applied: β-strand/β-bridge > α-helix > π-helix > 3_{10}-helix > hydrogen bonded turn > bend.

Results and discussion

Figure 2 shows the radius of gyration of all seven sets of protein configurations as described in Table 1 and Fig. 1. The NMR model structures as well as the simulations in water and TFE/water at pH 7 or pH 2 and 310 K (NMR_pH7_H_2O, MD_pH7_H_2O, MD_pH2_H_2O and MD_pH2_TFE) all have a radius of gyration around 1.4 nm: the NMR model structures at pH 7 are slightly more compact and the simulation in TFE slightly less compact than the simulation in water, while at pH 2 in water the protein starts unfolding after about 19 ns. As expected, at high temperature (HT, see Table 1) the protein becomes less compact due to partial unfolding with increased sampling of protein configurations during the heating procedure, either during the indicated simulation, MD_pH2_TFE_DR, or in the previous one, as is the case for MD_pH2_TFE_HT_DR. Both sets of protein configurations reach a maximum radius of gyration of about 2 nm but do not shrink back to 1.4 nm during or after the cooling sequence. Since the configuration of HEWL in water at pH 7 is clearly different (Buck et al. 1995) from the ones in TFE/water at pH 2, it is not surprising that the simulations with applied distance restraints derived from experimental NMR NOE proton–proton distances of HEWL in TFE/water at pH 2 show a different, increased radius of gyration with proceeding simulation time too. The simulation at 310 K using distance restraints (MD_pH2_TFE_DR) reaches even larger values for the radius of gyration than the free MD_pH2_TFE_HT simulation. This indicates that the water state and TFE state of HEWL are structurally different or, at least, a sizeable change in the protein size is required to get from one state to the other. This increased radius of gyration in TFE/water is also observed in small-angle X-ray scattering data (Hoshino et al. 1997) for HEWL.

Additional information about the observed structural changes can be gained from the secondary structure analysis (see Figs. 3, 4 and 5). The two simulations at pH 2 in water and TFE/water as well as the simulation in water at pH 7 (MD_pH2_H_2O, MD_pH2_TFE, and MD_pH7_

Fig. 4 Secondary structure elements (Kabsch and Sander 1983) as a function of time calculated for (from top to bottom) the MD_pH7_H$_2$O, MD_pH2_H$_2$O and MD_pH2_TFE simulation trajectories. *Red* α-helix; *green* π-helix; *black* 3$_{10}$-helix; *blue* β-strand; *yellow* β-bridge; *brown* bend; *grey* turn

H$_2$O, Fig. 4) all show the characteristic helices and β-strands in agreement with the secondary structure analysis of the NMR model structures (compare Fig. 3). A transition from the native state of HEWL, as it is present in aqueous solution at pH 7, to the TFE state at pH 2 reported by Buck et al. (1993, 1995), especially the transformation of the β-sheet region (residues 42–60) to an α-helix for residues 50–58 (compare Figure 9 a/b of Buck et al. (1995)), is not seen in any of the three simulations mentioned above, although in MD_pH2_H$_2$O the first β-strand is only intermittently present and the second one is shorter. Yet, a formation of a helix within the same residue range is not observed, even if the propensity for α-helices seems to be higher in TFE/water at pH 2 than in water only, in particular at pH 7. The picture is different when enabling the expected β-strand-to-α-helix transition using a temporally increased temperature or distance restraints based on experimentally derived NMR NOE proton–proton distance bounds of the TFE state of HEWL (MD_pH2_TFE_HT and MD_pH2_TFE_DR, Fig. 5). The heating from 310 to 400 K results in the complete loss of β-strands within residues 42–60 after about 10 ns, which happens even

faster, within 0.5 ns, when applying NOE distance restraints. The heating procedure not only affects the β-domain of the protein but also leads to a temporary loss of the two helices within residues 110–125. However, the protein structure is not that much disrupted to inhibit a reformation of these two helices in the continuation of the simulation including the cooling procedure. Application of distance restraining dissolves the α-helices around residues 30 and 110 and does not allow reformation of these helices. The loss of secondary structure is even larger when the heating/cooling procedure is followed by distance restraining (MD_pH2_TFE_HT_DR, Fig. 5), where all but two helices disappear and the transition of the β-strand region (residues 42–60) to an α-helix is not observed. This illustrates that the use of high temperature in an MD simulation strongly enhances the sampling, but it may be enhanced in the wrong or non-relevant direction in the high-dimensional conformational space, thereby leading to worse results than at low temperature.

Table 3 gives a quantitative overview of the positional root-mean-square difference between pairs out of five sets of protein configurations generated by MD simulation. The

Fig. 5 Secondary structure elements (Kabsch and Sander 1983) as a function of time calculated for (from top to bottom) the MD_pH2_TFE_HT, MD_pH2_TFE_DR and MD_pH2_TFE_HT_DR simulation trajectories. *Red* α-helix; *green* π-helix; *black* 3₁₀-helix; *blue* β-strand; *yellow* β-bridge; *brown* bend; *grey* turn
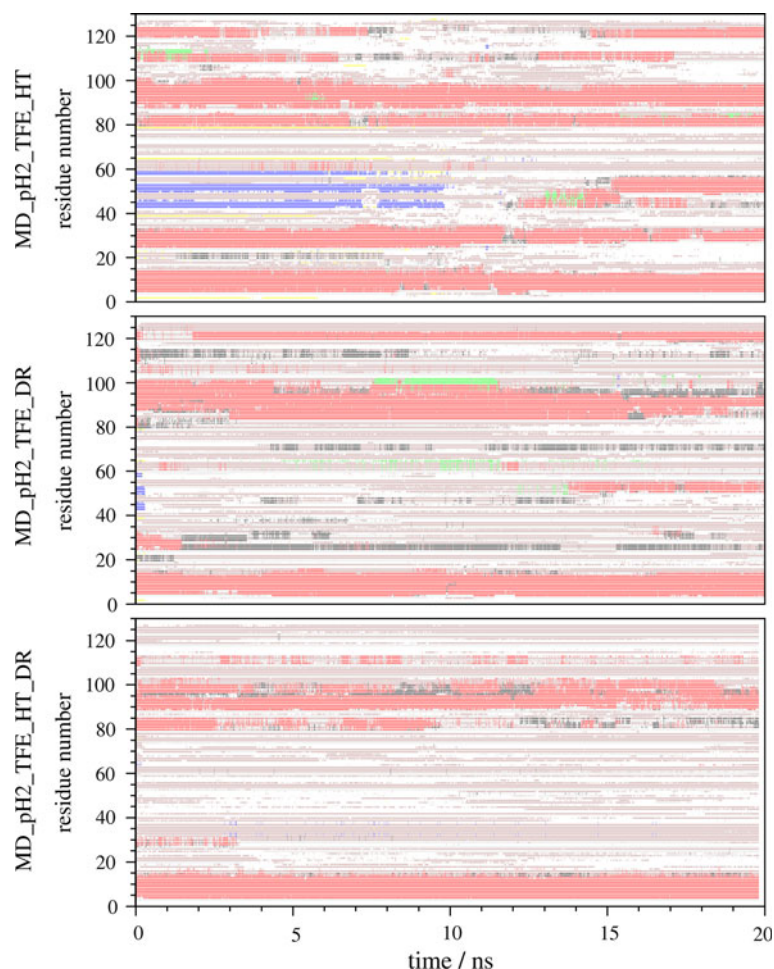

Table 3 Atom-positional root-mean-square deviation from the final configuration of one MD simulation to the final configuration of another MD simulation
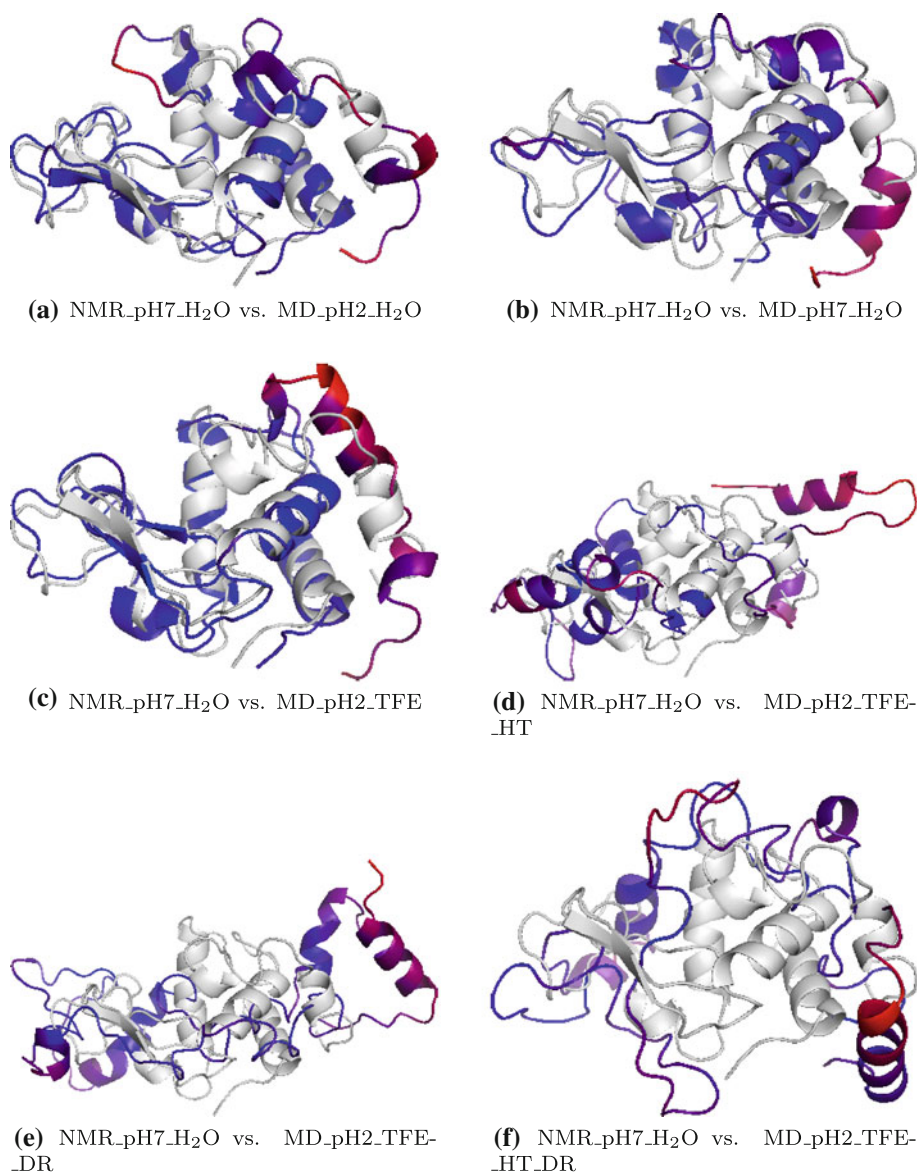
	MD_pH2_H$_2$O	MD_pH2_TFE	MD_pH2_TFE_HT	MD_pH2_TFE_DR	MD_pH2_TFE_HT_DR
MD_pH2_H$_2$O	0.00	0.41	1.43	1.68	1.40
MD_pH2_TFE	0.50	0.00	1.48	1.68	1.36
MD_pH2_TFE_HT	1.54	1.61	0.00	1.08	1.55
MD_pH2_TFE_DR	1.79	1.80	1.21	0.00	1.64
MD_pH2_TFE_HT_DR	1.50	1.46	1.67	1.74	0.00

Upper-right triangle: values for the C$_\alpha$ atoms only. Lower-left triangle: values for all atoms of the protein

same information is qualitatively shown in Figs. 6, 7 and 8. As we have seen from the analysis of the radius of gyration and the secondary structure, the two simulations at pH 2 without heating or distance restraining show the smallest deviation from each other (0.5 nm, see also Fig. 7b) and from the NMR model structure, see Fig. 6a and c. The difference from one of these two configurations to one of the HT or DR configurations is much larger, 1.46–1.79 nm, depending on which HEWL structures are compared. Interestingly, the three HT or DR configurations have rmsd values ranging from 1.21 to 1.74 nm between each other,

underlining the highly visible difference between these structures, compare also Fig. 9b. It also indicates that the different approaches in the simulations to obtain a TFE state configuration do not lead to the same result. This may reflect the wide diversity of conformations populated in the dynamic ensemble of a partially folded protein. The mentioned numbers result all from the rmsd calculations considering all protein atoms (lower-left triangle of Table 3) but the overall picture does not change when basing the analysis on the C$_\alpha$-atoms only in the upper-right triangle of Table 3. Therefore, the differences are not solely due to

(a) NMR_pH7_H$_2$O vs. MD_pH2_H$_2$O

(b) NMR_pH7_H$_2$O vs. MD_pH7_H$_2$O

(c) NMR_pH7_H$_2$O vs. MD_pH2_TFE

(d) NMR_pH7_H$_2$O vs. MD_pH2_TFE-_HT

(e) NMR_pH7_H$_2$O vs. MD_pH2_TFE-_DR

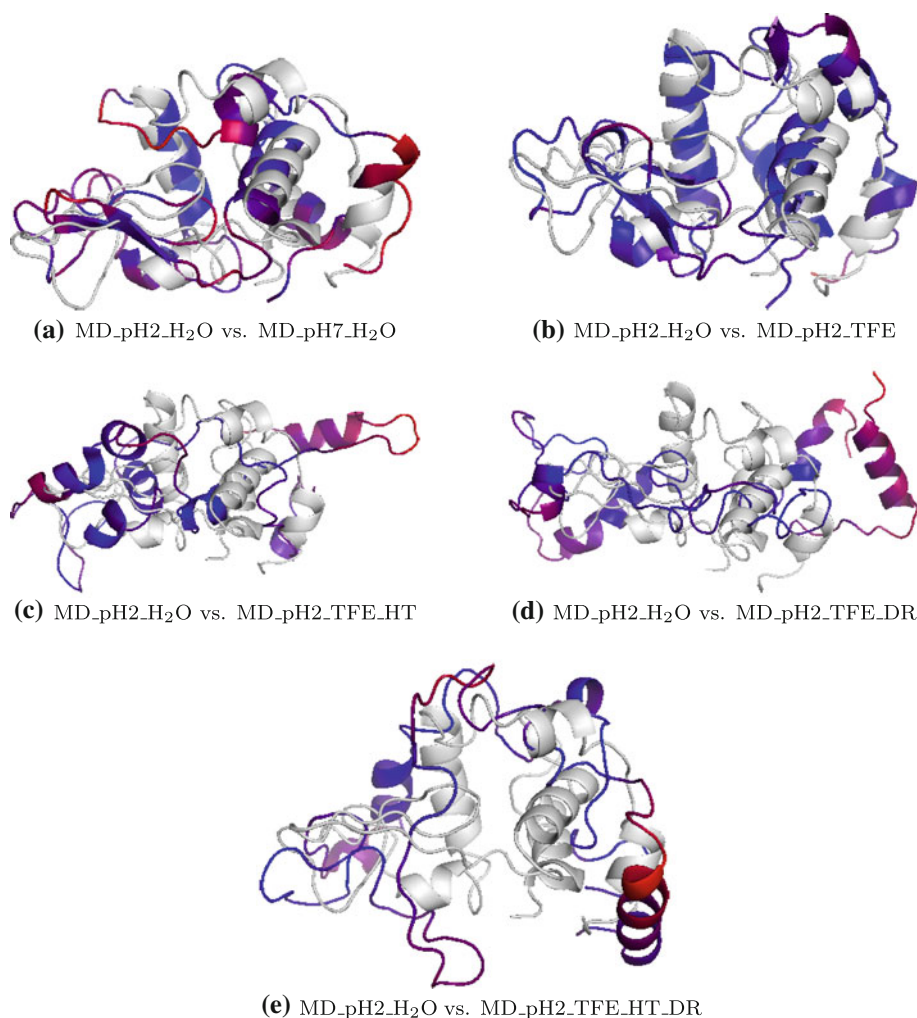(f) NMR_pH7_H$_2$O vs. MD_pH2_TFE-_HT_DR

different configurations of the amino acid side chains in pure water or TFE/water, but must result from conformational changes of the protein backbone.

A further characterisation of the different sets of configurations of HEWL at pH 2 in TFE can be obtained using the corresponding NOE data. NOE proton–proton upper bound distances were calculated from all seven sets of HEWL protein configurations (see Table 1) and compared to six NMR NOE proton–proton distance bound sets derived by experiment, as defined in Table 2. The resulting NOE bound violations are listed in Table 4. The protein configurations corresponding to the native state (NMR_pH7_H$_2$O, MD_pH7_H$_2$O) show the least violations of the H$_2$O NOE bound set and show much more violations with respect to the TFE NOE bound set. For the protein configurations corresponding to the TFE state, MD_pH2_TFE,

MD_pH2_TFE_HT, MD_pH2_TFE_DR, and MD_pH2_HT_DR, only for the latter three the opposite is the case. The three unrestrained simulations at 310 K (MD_pH7_H$_2$O, MD_pH2_H$_2$O and MD_pH2_TFE) show that the change from pH 7 to pH 2 is raising the number of NOE bound violations with respect to the H$_2$O NOE bounds set while only slightly reducing the number violations of the TFE NOE bound set. The secondary structure analysis showed that in MD_pH2_TFE_HT the three β-strands around residues 42–60 are dissolved and α-helical configurations emerged. This is reflected in a significant reduction of the violations involving residues 42–60 of the TFE NOE bound set compared to the simulation MD_pH2_TFE. However, the large number of NOE bound violations in regard to the NOE TFE set indicate that the configurations of MD_pH2_TFE_HT are not representative for the TFE state of the protein. The sets of

Fig. 7 Final configurations of two MD simulations aligned and colored according to the positional root-mean-square deviaton of the C_α atoms. *Dark blue* is good alignment, larger deviations are in *red*. The simulation names are defined in Table 1
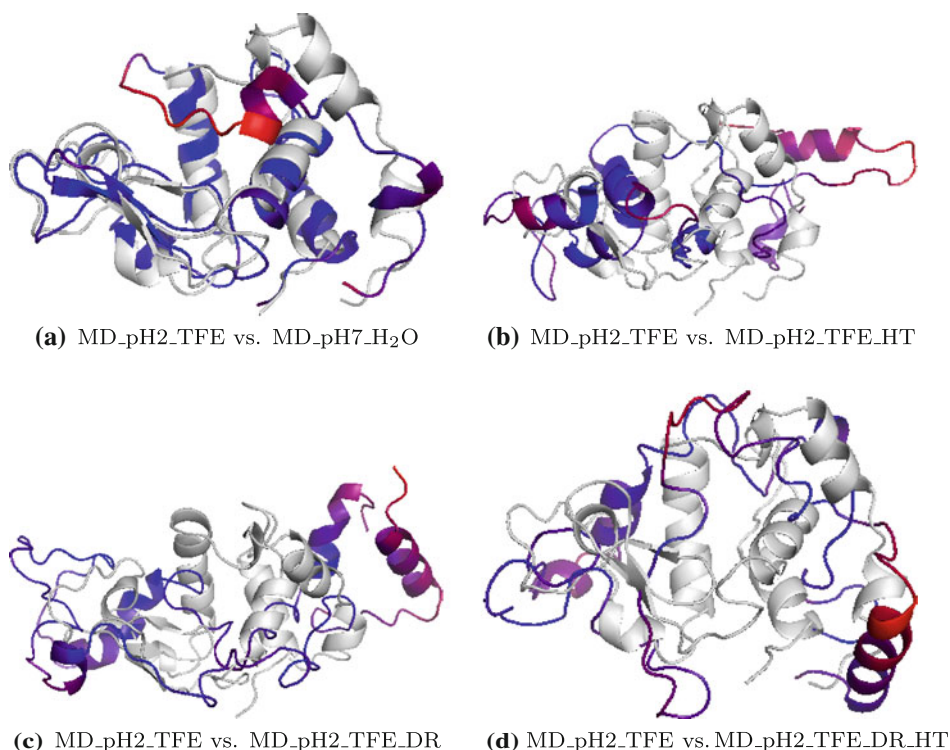

(a) MD_pH2_H$_2$O vs. MD_pH7_H$_2$O

(b) MD_pH2_H$_2$O vs. MD_pH2_TFE

(c) MD_pH2_H$_2$O vs. MD_pH2_TFE_HT

(d) MD_pH2_H$_2$O vs. MD_pH2_TFE_DR

(e) MD_pH2_H$_2$O vs. MD_pH2_TFE_HT_DR

configurations generated using NOE distance restraints for the TFE state do satisfy the NMR NOE TFE set of proton–proton distance bounds well, as expected, and as such both represent to some extent the configurational ensemble of the TFE state. Compared to the total number of NOE distance bounds in the two experimental NOE data sets the number of violations of the NOE TFE set for the DR configurations is even smaller than in the simulation of native HEWL in pure water at pH 7 with respect to the NOE_H$_2$O bound set (MD_pH7_H$_2$O, MD_pH2_TFE_DR, MD_pH2_TFE_HT_DR: 5, 1, 2 %). This does still not mean that we have found a unique configurational ensemble representing the TFE state of HEWL. Figure 9b and Table 3 show an rmsd of 1.74 nm between the two final configurations of MD_pH2_TFE_DR and MD_pH2_TFE_HT_DR. The number of 1388 experimental NMR NOE proton–proton upper bound distances in the NOE_TFE set is far too low to fully determine the structure of a protein with as many degrees of freedom as HEWL.

Figure 10 unifies the analysis of the NOE proton–proton distance bound violations with respect to both NOE bound

sets NOE_H$_2$O and NOE_TFE with the analysis of the secondary structure of the seven sets of protein configurations. The NMR model structures of HEWL at pH 7 in water do satisfy the NOE_H$_2$O bound set but not the NOE_TFE bound set for the residue range 35–66 and 100–129. The simulation of HEWL at pH 7 in water shows some violations of the NOE_H$_2$O bound set, but reproduces the secondary structure, while showing violations of the NOE_TFE set for about the same residue ranges as the NMR model structures. Applying NOE_TFE distance restraints the secondary structure is partly lost, helix B is shortened or gone and the β-strands are lost, while the NOE_TFE bounds are satisfied and the NOE_H$_2$O bounds are violated throughout the whole residue range.

The results from different analysis methods described above indicate that the two sets of DR configurations do satisfy the experimental data for the TFE state of HEWL. However, it is hard to decide which configuration set is the better one, since the various analysis methods do not yield a unique picture. Table 5 lists the protein-protein and protein-solvent potential energies for the seven sets of

(a) MD_pH2_TFE vs. MD_pH7_H$_2$O

(b) MD_pH2_TFE vs. MD_pH2_TFE_HT

(c) MD_pH2_TFE vs. MD_pH2_TFE_DR

(d) MD_pH2_TFE vs. MD_pH2_TFE_DR_HT

(a) MD_pH2_TFE_DR vs. MD_pH7_H$_2$O

(b) MD_pH2_TFE_DR vs. MD_pH2_TFE_HT_DR

configurations. At pH 7 MD simulation of HEWL in aqueous solution slightly lowers the solute-solute energy compared to the set of NMR model structures. Lowering the pH from 7 to 2 results in an increase of the solute-solute energy which is due to the increase of the total charge of the solute from +8 e to +19 e. Considering the four trajectories of HEWL in TFE/water, MD_pH2_TFE has solute-solute and solute-solvent energies closest to MD_pH2_H$_2$O. Raising the temperature leads to an increase in solute-solute energy. At higher temperature, the influence of the solvent is enhanced. Interestingly, the application of NOE distance restraints yields the lowest solute-solute energy and total energy of the protein. The restraints apparently focuses the sampling on energetically more favourable conformations of the protein. Comparing both distance restrained trajectories, the one at room temperature, MD_pH2_TFE_DR, yields the lowest energies. So

this conformational ensemble offers the best representation of the TFE state of HEWL.

A comparison of calculated with measured (Buck et al. 1993) values of observables accessible by spectroscopic techniques such as far-UV CD or fluorescence spectroscopy for HEWL (Buck et al. 1993; Povey et al. 2007; D'Amico et al. 2011) or its constituting peptides (Yang et al. 1995; Povey et al. 2007) would not be very conclusive for the following reasons: (i) CD spectra only indicate overall helical content, not which helices are present or not; (ii) fluorescence data only reflect modifications in the relative positions of the six Trp side chains of HEWL, but not the details with respect to the individual Trp residues; (iii) our MD simulations were done in 70 % TFE in order to match the conditions of the NMR experiment (Buck et al. 1995), whereas the CD and fluorescence measurements were done at 15 % (Buck et al.

Table 4 Number of NOE bound violations of the seven sets of protein structures with respect to two experimental data sets of HEWL in water (Schwalbe et al. 2001) and HEWL in a TFE/water mixture (Buck 1994)

System	Solvent	Selection	Number of NOE violations		
			>0.1 nm	>0.3 nm	>0.5 nm
NMR_pH7_H₂O	Water	All	0/0/0/0	0/0/0/0	0/0/0/0
		Only	0/0/0/0	0/0/0/0	0/0/0/0
		Residues 42–60	0/0/0/0	0/0/0/0	0/0/0/0
	TFE/water	All	167/50/111/6	70/15/49/6	29/3/22/4
		Only	167/50/111/6	70/15/49/6	29/3/22/4
		Residues 42–60	22/7/15/0	8/2/6/0	3/1/2/0
MD_pH7_H₂O	Water	All	89/12/28/49	27/1/5/21	13/0/2/11
		Only	84/9/26/49	27/1/5/21	13/0/2/11
		Residues 42–60	2/0/1/1	0/0/0/0	0/0/0/0
	TFE/water	All	199/48/143/8	80/14/60/6	30/2/22/6
		Only	199/48/143/8	80/14/60/6	30/2/22/6
		Residues 42–60	23/6/17/0	7/2/5/0	1/0/1/0
MD_pH2_H₂O	Water	All	124/9/21/94	45/1/1/43	21/0/0/21
		Only	121/7/20/94	45/1/1/43	21/0/0/21
		Residues 42–60	10/0/4/6	1/0/0/1	0/0/0/0
	TFE/water	All	179/51/121/7	75/14/54/7	23/0/18/5
		Only	179/51/121/7	75/14/54/7	23/0/18/5
		Residues 42–60	25/8/17/0	10/3/7/0	4/0/4/0
MD_pH2_TFE	Water	All	101/9/17/75	51/1/4/46	29/0/0/29
		Only	98/6/17/75	50/0/4/46	29/0/0/29
		Residues 42–60	3/0/3/0	0/0/0/0	0/0/0/0
	TFE/water	All	180/48/123/9	72/10/56/6	26/0/22/4
		Only	177/47/121/9	72/10/56/6	26/0/22/4
		Residues 42–60	27/7/20/0	9/3/6/0	3/0/3/0
MD_pH2_TFE_HT	Water	All	448/32/89/327	344/2/38/304	294/0/15/279
		Only	428/23/78/327	342/2/36/304	293/0/14/279
		Residues 42–60	42/2/9/31	31/0/5/26	24/0/1/23
	TFE/water	All	193/41/143/9	72/5/60/7	34/2/26/6
		Only	188/41/138/9	71/5/59/7	34/2/26/6
		Residues 42–60	13/8/5/0	1/0/1/0	0/0/0/0
MD_pH2_TFE_DR	Water	All	387/20/61/306	268/0/16/252	221/0/3/218
		Only	378/14/58/306	268/0/16/252	221/0/3/218
		Residues 42–60	41/1/9/31	29/0/4/25	23/0/1/22
	TFE/water	All	11/7/4/0	0/0/0/0	0/0/0/0
		Only	11/7/4/0	0/0/0/0	0/0/0/0
		Residues 42–60	0/0/0/0	0/0/0/0	0/0/0/0
MD_pH2_TFE_HT_DR	Water	All	405/29/71/305	291/1/21/269	247/0/7/240
		Only	385/18/62/305	289/0/20/269	246/0/6/240
		Residues 42–60	44/4/11/29	29/0/4/25	22/0/1/21
	TFE/water	All	21/10/11/0	1/0/1/0	0/0/0/0
		Only	17/9/8/0	0/0/0/0	0/0/0/0
		Residues 42–60	1/1/0/0	0/0/0/0	0/0/0/0

"/" separate all/short/medium/long ranged NOE bound distances as defined in Table 2. Each experimental data set is further split into "all" (containing the NOEs of the whole data set), "only" (containing the NOEs which occur in the corresponding set only) and "residues 42–60" (containing NOEs which involve atoms of residues 42–60)
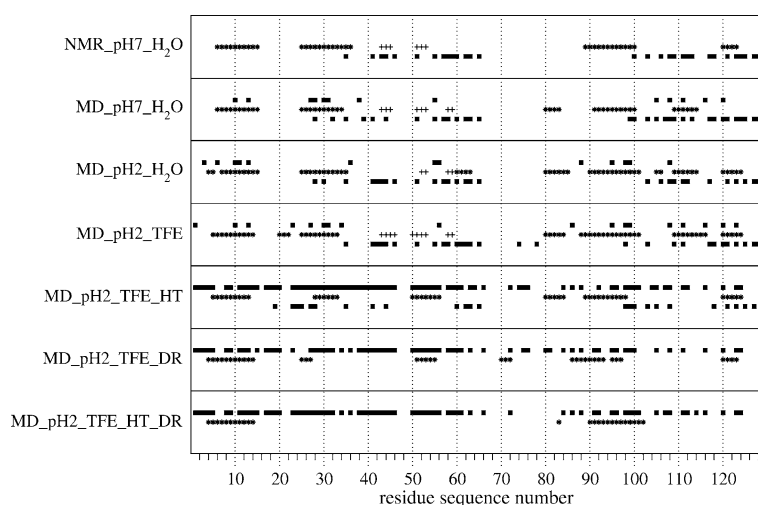
Fig. 10 NOE distance bound violations larger than 0.5 nm for atom pairs with their position within the protein residue sequence for the seven different sets of structures averaged over the final 4 ns of the simulation trajectories. Each system shows the secondary structure (*: helix, +: β-strand) of the corresponding residue. Residues with one ore more atoms involved in a violation of the experimental NOE sets are marked *black*. Violations of the experimental NOE bound set NOE_H₂O of HEWL in water are above the seconday structure, while those of the NOE_TFE set are below the secondary structure

Table 5 Protein-protein and protein-solvent potential energies for the seven sets of HEWL configurations

| System	Potential energy/kJ mol⁻¹				
	Solute-solute			Solute-solvent	Total
	Bonded	Nonbonded	Total		
NMR_pH7_H₂O	4,312	−11,290	−6,978	–	–
MD_pH7_H₂O	3,855	−11,152	−7,297	−12,889	−20,186
MD_pH2_H₂O	4,454	−9,463	−5,009	−12,204	−17,213
MD_pH2_TFE	4,396	−9,734	−5,338	−11,420	−16,758
MD_pH2_TFE_HT	4,476	−8,405	−3,929	−14,064	−17,993
MD_pH2_TFE_DR	5,108	−11,181	−6,073	−13,329	−19,402
MD_pH2_TFE_HT_DR	5,167	−10,948	−5,781	−13,264	−19,045

The energies calculated from MD simulation trajectories are averaged over the final 5 ns, while for NMR_pH7_H₂O the average is calculated over the 50 NMR model configurations

1993), 50 % (Yang et al. 1995), 40 % (Hoshino et al. 1997), 0–50 % (Povey et al. 2007) and 0–20 % (D'Amico et al. 2011) TFE.

Figure 11 shows the radial distribution functions from the centre of geometry of all protein atoms to the oxygen atoms of the solvent molecules for the six sets of HEWL configurations generated by MD simulations. There are almost no water molecules closer than 1 nm to the centre of the protein in pure water at pH 7. Lowering the pH to 2, the protein starts to unfold after 19 ns of simulation time (compare Fig. 2) and water molecules penetrate into the protein. A native-like fold in TFE/water at pH 2 (MD_pH2_TFE) seems to prefer protein-TFE interaction over protein-water interaction as indicated by the peak of

the first solvation layer of TFE around 0.5 nm. This is in agreement with results from previous MD simulations of peptides in explicit TFE/water solutions which showed an accumulation of TFE molecules at the peptide surface (Fioroni et al. 2002; Roccatano et al. 2002; Diaz et al. 2002; Mehrnejad et al. 2007). At high temperature or with distance restraints applied (MD_pH2_TFE_HT and MD_pH2_TFE_DR) both solvents penetrate the protein. MD_pH2_TFE_DR_HT allows the highest solvent penetration, a result of the less compact and more unfolded protein configuration in this ensemble. As mentioned before, a detailed analysis of protein-TFE interactions and structural properties lies beyond the scope of the present study.

Fig. 11 Radial distribution functions, $g(r)$, calculated from the final 5 ns of the trajectories of the six different MD simulations of HEWL. The protein-TFE (*dotted lines*) and protein-water (*black lines*) distances are the distances between the centre of geometry of all HEWL atoms and the TFE oxygen atoms or water oxygen atoms, respectively
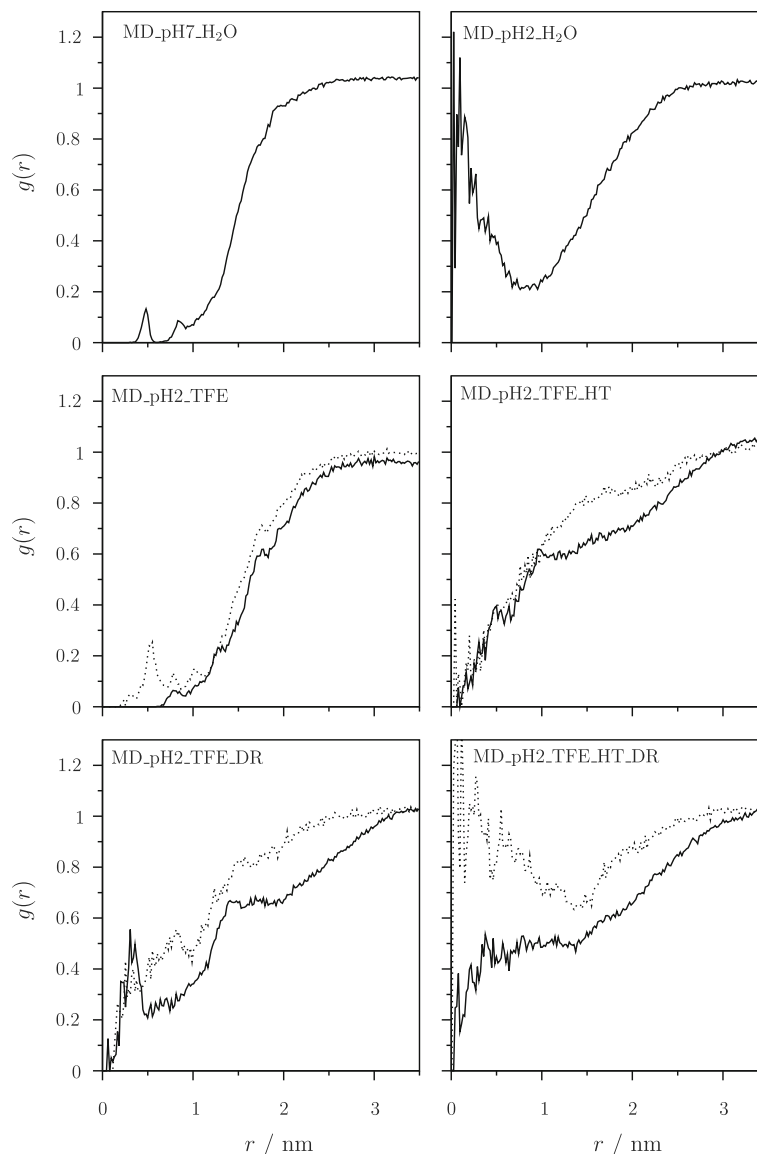
Conclusion

Using a set of 1388 NOE atom-atom distance bounds measured for HEWL at pH 2 in 70 % TFE/30 % water it was attempted to generate a configurational ensemble for this protein in its TFE state which would confirm the conversion of the β-sheet between residues 42 and 60 present in HEWL at pH 7 in water into an α-helical segment of residues 50 to 58 in the TFE state.

Starting from a structure of HEWL at pH 7 in water and just changing the pH to 2 and the solvent to 70 % TFE/30 % water (called TFE) no major conformational changes were observed within 20 ns of MD simulation. The use of higher temperatures, up to 400 K, to enhance the sampling led to conformational changes and in particular cases to the emergence of α-helical structures for residues 50–58 that had β-strand secondary structure at pH 7 in water, but the

generated configuration did not satisfy the experimentally derived NOE distance upper bounds for HEWL in TFE/water. This could only be achieved by applying the set NOE_TFE of distance bounds as distance restraints in the MD simulations. The resulting sets of HEWL configurations do represent to some degree the configurational ensemble of the TFE state, but is by no means unique or complete, which is due to the limited number of NOE bounds, their distribution over the protein, to the limited sampling of configurational space within 20 ns of MD simulation and the wide conformational diversity expected in a partially folded state of this type. Yet, it represents the currently best possible representation of this ensemble.

References

http://www.gromos.net. URL http://www.gromos.net

Artymiuk PJ, Blake CCF, Rice DW, Wilson KS (1982) The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 Å resolution. Acta Crystallogr Sect B Struct Sci 38:778–783

Bégué JP, Bonnet-Delpon D, Crousse B (2004) Fluorinated alcohols: a new medium for selective and clean reaction. Synlett 1:18–29

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) chapter interaction models for water in relation to protein hydration. Reidel, Dordrecht, pp 331–342

Blanco FJ, Jimenez MA, Pineda A, Rico M, Santoro J, Nieto JL (1994) NMR solution structure of the isolated N-terminal fragment of protein-G b_1 domain: evidence of trifluoroethanol induced native-like β-hairpin formation. Biochemistry 33(19):6004–6014. doi:10.1021/bi00185a041

Buck M (1994) NMR Studies of the Dynamics and Folding of Hen Lyzozyme. PhD thesis, University of Oxford

Buck M (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. Q Rev Biophys 31(3):297–355

Buck M, Radford SE, Dobson CM (1993) A partially folded state of hen egg-white lysozyme in trifluoroethanol: structural characterization and implications for protein folding. Biochemistry 32(2):669–678. doi:10.1021/bi00053a036

Buck M, Schwalbe H, Dobson CM (1995) Characterization of conformational preferences in a partly folded protein by heteronuclear NMR-spectroscopy: assignment and secondary structure-analysis of hen egg-white lysozyme in trifluoroethanol. Biochemistry 34(40):13219–13232

Buck M, Schwalbe H, Dobson CM (1996) Main-chain dynamics of a partially folded protein: N-15 NMR relaxation measurements of hen egg white lysozyme denatured in trifluoroethanol. J Mol Biol 257(3):669–683

Cammers Goodwin A, Allen TJ, Oslick SL, McClure KF, Lee JH, Kemp DS (1996) Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol. J Am Chem Soc 118(13):3082–3090. doi:10.1021/ja952900z

Chiti F, Taddei N, Bucciantini M, White P, Ramponi G, Dobson CM (2000) Mutational analysis of the propensity for amyloid formation by a globular protein. EMBO J 19(7):1441–1449. doi:10.1093/emboj/19.7.1441

D'Amico M, Raccosta S, Cannas M, Martorana V, Manno M (2011) Existence of metastable intermediate lysozyme conformation highlights the role of alcohols in altering protein stability. J Phys Chem B 115(14):4078–4087. doi:10.1021/jp106748g

Diaz MD, Fioroni M, Burger K, Berger S (2002) Evidence of complete hydrophobic coating of bombesin by trifluoroethanol in aqueous solution: an NMR spectroscopic and molecular dynamics study. Chem Eur J 8(7):1663–1669. doi:10.1002/1521-3765(20020402)8:7<1663

Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang DQ,

van Gunsteren WF (2011) GROMOS++ software for the analysis of biomolecular simulation trajectories. J Chem Theory Comput 7(10):3379–3390. doi:10.1021/ct2003622

Fezoui Y, Teplow DB (2002) Kinetic studies of amyloid β-protein fibril assembly: differential effects of α-helix stabilization. J Biol Chem 277(40):36948–36954. doi:10.1074/jbc.M204168200

Fioroni M, Burger K, Mark AE, Roccatano D (2000) A new 2,2,2-trifluoroethanol model for molecular dynamics simulations. J Phys Chem B 104(51):12347–12354. doi:10.1021/jp002115v

Fioroni M, Diaz MD, Burger K, Berger S (2002) Solvation phenomena of a tetrapeptide in water/trifluoroethanol and water/ethanol mixtures: a diffusion NMR, intermolecular NOE, and molecular dynamics study. J Am Chem Soc 124(26):7737–7744. doi:10.1021/ja0259335

Harvey SC, Tan RKZ, Cheatham TE (1998) The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. J Comput Chem 19(7):726–740 ISSN 01928651

Heinz TN, van Gunsteren WF, Hünenberger PH (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. J Chem Phys 115(3):1125–1136

Hockney RW, Eastwood JW (1981) Computer simulation using particles. McGraw-Hill, New York

Hong DP, Hoshino M, Kuboi R, Goto Y (1999) Clustering of fluorine-substituted alcohols as a factor responsible for their marked effects on proteins and peptides. J Am Chem Soc 121(37):8427–8433

Hoshino M, Hagihara Y, Hamada D, Kataoka M, Goto Y (1997) Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering. FEBS Lett 416(1):72–76

Jasanoff A, Fersht AR (1994) Quantitative-determination of helical propensities from trifluoroethanol titration curves. Biochemistry 33(8):2129–2135. doi:10.1021/bi00174a020

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637

Kumar S, Modig K, Halle B (2003) Trifluoroethanol-induced $\beta \rightarrow \alpha$ transition in β-lactoglobulin: Hydration and cosolvent binding studied by ^{2}H, ^{17}O, and ^{19}F magnetic relaxation dispersion. Biochemistry 42(46):13708–13716. doi:10.1021/bi035330l

Lu H, Buck M, Radford SE, Dobson CM (1997) Acceleration of the folding of hen lysozyme by trifluoroethanol. J Mol Biol 265(2):112–117. doi:10.1006/jmbi.1996.0715

Mehrnejad F, Naderi-Manesh H, Ranjbar B (2007) The structural properties of magainin in water, TFE/water, and aqueous urea solutions: molecular dynamics simulations. Proteins 67(4):931–940. doi:10.1002/prot.21293

Nelson JW, Kallenbach NR (1986) Stabilization of the ribonuclease S-peptide α-helix by trifluoroethanol. Proteins Struct Funct Genet 1(3):211–217. doi:10.1002/prot.340010303

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25(13):1656–1676. doi:10.1002/jcc.20090

Povey JF, Smales CM, Hassard SJ, Howard MJ (2007) Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function. J Struct Biol 157(2):329–338. doi:10.1016/j.jsb.2006.07.008

Rajan R, Balaram P (1996) A model for the interaction of trifluoroethanol with peptides and proteins. Int J Pept Protein Res 48(4):328–336

Roccatano D, Colombo G, Fioroni M, Mark AE (2002) Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize

secondary-structure formation in peptides: a molecular dynamics study. Proc Natl Acad Sci USA 99(19):12179–12184. doi:10.1073/pnas.182199699

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341

Schmid N, Allison JR, Dolenc J, Eichenberger AP, Kunz APE, van Gunsteren WF (2011) Biomolecular structure refinement using the GROMOS simulation software. J Biomol NMR 51(3):265–281. doi:10.1007/s10858-011-9534-0

Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF (2012) Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. Comput Phys Commun 183(4):890–903. doi:10.1016/j.cpc.2011.12.014

Schwalbe H, Grimshaw SB, Spencer A, Buck M, Boyd J, Dobson CM, Redfield C, Smith LJ (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. Protein Sci 10(4):677–688

Shiraki K, Nishikawa K, Goto Y (1995) Trifluoroethanol-induced stabilization of the α-helical structure of β-lactoglobulin - implication for non-hierarchical protein-folding. J Mol Biol 245(2):180–194. doi:10.1006/jmbi.1994.0015

Shuklov IA, Dubrovina NV, Boerner A (2007) Fluorinated alcohols as solvents, cosolvents and additives in homogeneous catalysis. Synthesis (Stuttgart) 19:2925–2943

Smith LJ, Alexandrescu AT, Pitkeathly M, Dobson CM (1994) Solution structure of a peptide fragment of human α-lactalbumin in trifluoroethanol: a model for local-structure in the molten globule. Structure 2(8):703–712. doi:10.1016/S0969-2126(00)00071-X

Starzyk A, Barber-Armstrong W, Sridharan M, Decatur SM (2005) Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: an isotope-edited FTIR study. Biochemistry 44(1):369–376. doi:10.1021/bi0481444

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v., Zürich, Groningen

Wüthrich K, Billeter M, Braun W (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. J Mol Biol 169(4):949–961

Yang JJ, Buck M, Pitkeathly M, Kotik M, Haynie DT, Dobson CM, Radford SE (1995) Conformational properties of 4 peptides spanning the sequence of hen lysozyme. J Mol Biol 252(4):483–491. doi:10.1006/jmbi.1995.0513