Molecular Dynamics Simulation of Ester-Linked Hen Egg White Lysozyme Reveals the Effect of Missing Backbone Hydrogen Bond Donors on the Protein Structure

by **Andreas P. Eichenberger**, **Zrinka Gattin**, **Garif Yalak**, and **Wilfred F. van Gunsteren***

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, CH-8093 Zürich (phone: +41-44-6325501; e-mail: wfvgn@igc.phys.chem.ethz.ch)

The three-dimensional structure of a protein is stabilized by a number of different atomic interactions. One of these is hydrogen bonding. Its influence on the spatial structure of the hen egg white lysozyme is investigated by replacing peptide bonds (except those of the two proline residues) by ester bonds. Molecular dynamics simulations of native and ester-linked lysozyme are compared with the native crystal structure and with NOE distance bounds derived from solution NMR experiments. The ester-linked protein shows a slight compaction while losing its native structure. However, it does not unfold completely. The structure remains compact due to its hydrophobic core and a changed network of hydrogen bonds involving side chains.

Introduction. – Since the first elucidation of the spatial structure of a protein, the desire to understand the particular interactions that hold a protein in its native fold has driven a great variety of investigations, both of experimental and of theoretical, and computational nature [1]. The issue is a complex one because different types of interactions play a role [2]: H-bonds between the amide (NH) and carbonyl (CO) atoms of the peptidic linkages, *Van der Waals* interactions between tightly packed nonpolar side-chain atoms in the protein interior, hydrophobic effects due to the aqueous solvent environment, and other (de)solvation interactions. Apart from these major interactions determining protein spatial structure, less important ones have been identified, such as salt bridges, H-bonds involving side-chain atoms, and disulfide bridges [2]. A systematic investigation of the contribution of each of these types of interactions to the protein structure would separately eliminate each of these interactions and observe the subsequent changes in protein structure. Such a dissecting approach is easier to execute computationally than experimentally, because a computational study is less bound by chemical or physical limits. But, even computationally, it would be difficult to eliminate the hydrophobic effects without interfering with other interactions. An interaction that can be both computationally and experimentally eliminated with a minimal interference with other interactions that determine protein structure is H-bonding involving the NH and CO atoms of the peptide linkages in the polypeptide chain. Such H-bonds are largely absent in α-depsipeptides, naturally occurring peptides in which particular amino acid residues are replaced by hydroxyl-acid residues [3], *e.g.*, (*S*)-3-hydroxybutanoic acid in the case of a glycine residue, thereby replacing some of the backbone amide linkages by ester linkages. Examples of α-depsipeptides are beauvericin, valinomycin, and aureobasidin A [4–6]. The amide and ester linkages both are planar due to electronic resonance and

prefer the *syn* orientation [7]. Their H-bonding properties are different though. While the amide (peptide) linkage may act as H-bond donor (amide H-atom) as well as H-bond acceptor (carbonyl O-atom), the ester linkage may exclusively function as H-bond acceptor. Furthermore, the amide carbonyl O-atom is a substantially stronger H-bond acceptor than the ester carbonyl O-atom [8][9]. Thus α-depsipeptides are analogues of α-peptides with reduced H-bonding capabilities.

Amide-to-ester backbone substitution has been studied experimentally for the PIN WW domain, a three-stranded 34-residue protein [10–13]. The folding kinetics and thermodynamics showed that the protein is most destabilized, when H-bonds that are enveloped by a hydrophobic cluster are perturbed.

In the present study, we compared the structural properties of a well-known 129-residue protein, hen egg white lysozyme (HEWL), and its depsi analogue. Out of the many possible combinations of ester-with-peptide linkages, $\sum_{k=1}^{126} \binom{126}{k}$, we chose the one with only ester linkages, in which all backbone H-bond donors are eliminated. This will maximize the structural effects. When experimental data on HEWL with less than 126 ester linkages would become available, the study presented here could easily be repeated for the particular ester-linked analogue of HEWL.

A previous 3.5 ns molecular dynamics (MD) simulation of native HEWL in aqueous solution [14] is compared to that of 10 ns in which all peptide bonds, except those involving the two proline residues, have been replaced by ester bonds. Various structural properties are computed, and the sizeable effects of missing backbone–backbone H-bonds clearly show up.

Results. – *General Structural Properties and Flexibility.* The root-mean-square deviations (rmsd) of the trajectory structures of native and ester-linked HEWL from the native X-ray structure are shown in *Fig. 1,a*. The values for the backbone and for all atoms in the native HEWL simulation slightly increase during the first 2 ns leveling off around 0.20 and 0.30 nm, respectively. The rmsd values of the ester-linked HEWL increase much faster over the first 5 ns of the simulation and then remain at values around 0.55 and 0.65 nm. This indicates that the ester-linked protein changes its spatial structure.

The convergence of the radius of gyration (rgyr), a measure of the compactness of the structure, calculated for the ester-linked HEWL simulation is also faster than that for the native HEWL simulation (*Fig. 1,b*) and seems to reach a lower value. The molecule is compacting rapidly. This is less clearly observed in the solvent-accessible surface area (sasa; *Fig. 1,c*) and protein volume (*Fig. 1,d*).

The appearance and disappearance of secondary-structure elements during the simulations of native HEWL and ester-linked HEWL is shown in *Fig. 2*. The three β-strands quickly disappear in ester-linked HEWL but show a comeback between 7.5 to 8.5 ns (residues 45, 46, and 50–53). The α-helices are preserved a bit longer but also disappear during the first 3 ns without returning later. In contrast, the π-helix (residues 111–115) survives during the whole simulation time of 10 ns, but appears to be structurally less stable than in native HEWL (occurrence in the X-ray structure/native HEWL/ester-linked HEWL: 100/71/43%). The loss of secondary structure is also illustrated in *Fig. 3* which shows different structures along the MD trajectories.

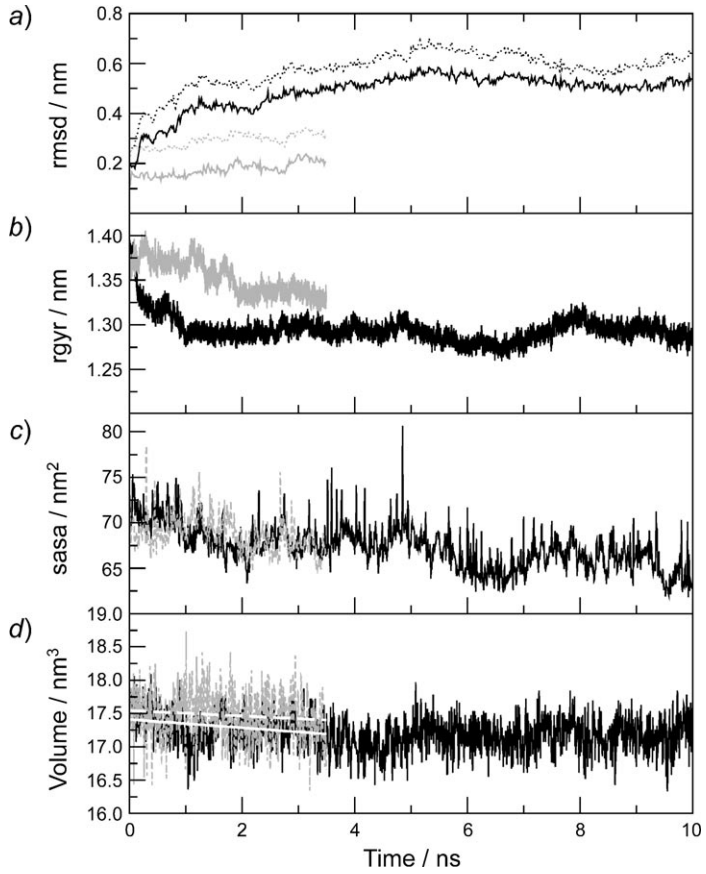

Fig. 1. a) *Root-mean-square deviations* (rmsd) *of native HEWL and ester-linked HEWL from the initial X-ray crystal structure.* The values are calculated for the whole simulation time (native/ester-linked HEWL: 3.5/10 ns) considering all atoms (dotted lines) or backbone atoms only (native: N, C, O, C_α; ester-linked: O_α, C, O C_α; solid lines). b) *Radius of gyration* (rgyr) *of backbone atoms* (native HEWL (gray): N, C, O, C_α ; ester-linked HEWL (black): O_α, C, O, C_α). The values are calculated through the whole simulation time of 3.5 ns (native HEWL) and 10 ns (ester-linked HEWL). c) *The solvent-accessible surface area* (sasa), *calculated for native HEWL* (0–3.5 ns; dashed gray line) *and ester-linked HEWL* (0–10 ns; solid black line). d) *Protein volume calculated for ester-linked HEWL* (solid, black) *and native HEWL* (dashed, gray line) *as well as a linear regression of the protein volume over the first 3.5 ns* (native/ester-linked HEWL: dashed/solid line).

The φ/ψ-angle distribution of residues 5–14, 25–36, 80–84, 89–101, 109–114, and 120–124 are shown in *Fig. 4*. These residues are the ones involved in an α-helix for at least 33% of the time during the 3.5 ns of the native HEWL simulation [17]. In the time window of 1.5–3.5 ns, 80% of all φ/ψ-angle combinations of the native HEWL simulation are in the region of an α-helix (*Fig. 4,b*). The ester-linked HEWL simulation shows a different φ/ψ-angle distribution during the same time period (*Fig. 4,c*). The distribution of the φ-angle is broader with its maximum value shifted from $-60°$ to $-90°$. This broadening effect does not occur for the ψ-angle distribution.

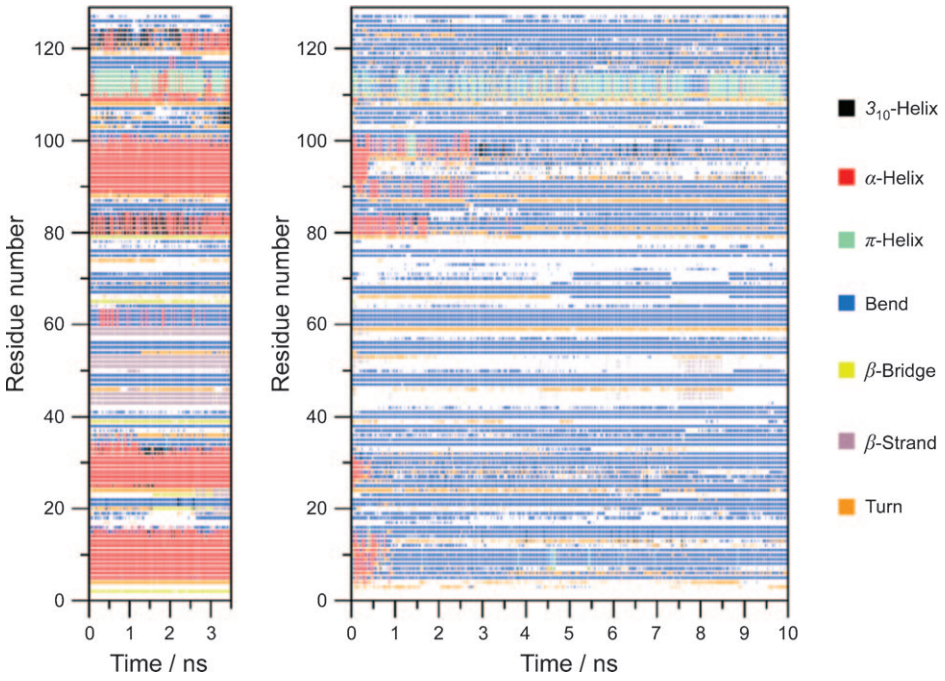

Fig. 2. *Secondary structure elements as a function of time calculated for the two MD simulation trajectories of native and ester-linked HEWL.* The missing H-atoms of the ester-linked HEWL MD simulation trajectory to analyze secondary structure elements were placed using the M procedure.

It becomes bimodal with a maximum at about the same position ($-60°$) and width as observed in the native HEWL simulation, but accompanied by a second maximum positioned at the inverse angle ($+60°$) of the global maximum. This picture does not change during the simulation (*Fig. 4*, *d*).

Fig. 5 shows the rms fluctuations of the C_α-atoms in the native and ester-linked HEWL simulations. Calculations were performed over two different time windows of the same length, one from 1.5–3.5 ns (last 2 ns of the native HEWL simulation), the other from 8–10 ns (last 2 ns of the ester-linked HEWL simulation). The values are not significantly different when comparing the native HEWL to the ester-linked HEWL during the first time window. The most substantial differences appear for residues 80–100 and 116–129. The rms fluctuations calculated for the second time window (8–10 ns) show locally much higher values than the native simulation (*e.g.*, around residue 68), but are in general quite comparable. This confirms that the ester-linked protein is not completely unfolding and stays relatively compact.

NOE Proton–Proton Distances. The differences in the structure between native and ester-linked HEWL can also be analyzed in terms of NOE atom–atom distance bound violations using the NOE bounds determined from NMR experiments on native HEWL [18]. The results are shown in *Fig. 6* and compiled in *Table 1*.

Two sets of NOE bounds were used. Set 1 includes all 1630 NOE bounds, while set 2 contains 392 bounds between atoms in residues i and j with $j \geq i+4$ only. The distance

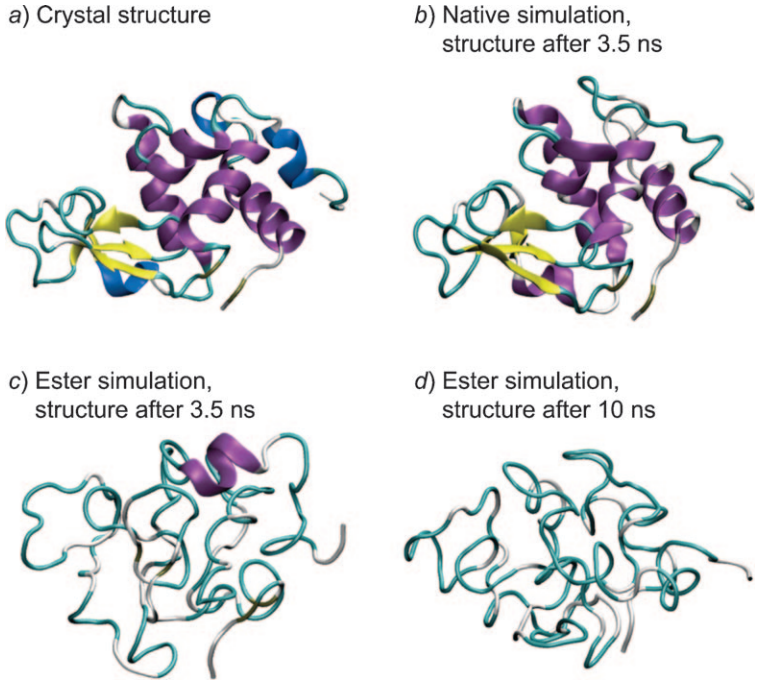

Fig. 3. *Schematic diagram showing the native HEWL crystal structure* (*a*) [15]*, structures after 3.5 ns simulations of native HEWL* (*b*) *and ester-linked HEWL* (*c*)*, and after 10 ns simulation of ester-linked HEWL* (*d*). Color represent: α-helix (purple), 3_{10}-helix (blue), β-strand (yellow), β-bridge (brown), turn (cyan), and coil (white), according to the rules implemented in VMD [16]. In the X-ray structure, the four helices (A: residues 5–14; B: residues 25–36; C: residues 89–101; D: residues 109–114) are in purple, the three 3_{10}-helices (residues 20–22, 80–84 and 120–124) in blue, and the three β-sheets (residues 43–45, 51–53 and 58–59) in yellow (*cf.* Table 1 of [17]).

bound violations were also calculated for the crystal structure [15]. The total number of violations is much higher in the ester-linked structure than in the X-ray structure and most of the violations are observed for the long-range NOE bounds (set 2). Yet, there are only twelve violations larger than 0.5 nm. These involve only ten residues (12, 13, 17, 28, 88, 95, 98, 107, 108, and 129) and seven pairs of residues. Of theses seven pairs, only two pairs (12–28 and 98–107) connect two secondary structure elements (helices A–B and C–3_{10}), the other five pairs connect helices (A, B, and C) with loops. Thus, less than 1% of the number of NOE bonds is violated by more than 0.5 nm. Or, 85% of all 1630 bounds are satisfied within 0.1 nm. Since the structure of ester-linked HEWL differs significantly from the native one (see *Figs. 1* and *3*), this means that a large part of the set of NOE bounds holds little spatial information regarding HEWL.

Hydrogen Bonding. Because the ester-linked protein is not unfolding, one may ask whether intramolecular H-bonding is decreased upon replacing peptide NH groups by ester O_α atoms. Table S1 (*Supplementary material*)[1]) lists the occurrence of the H-

[1]) Available from the authors upon request.

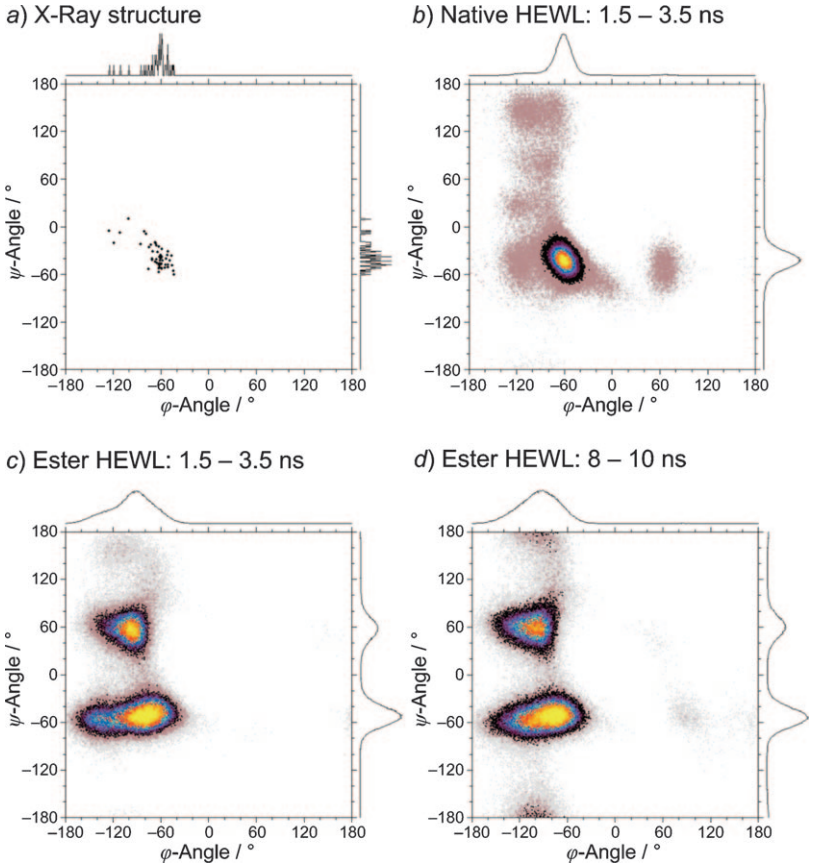

Fig. 4. Ramachandran *plots based on* a) *the X-ray structure of native HEWL* [15], b) *the simulation trajectory of native HEWL from 1.5–3.5 ns, and* c) – d) *the simulation trajectory of ester-linked HEWL calculated through time windows of 1.5–3.5 and 8–10 ns, respectively.* The one-dimensional φ- and ψ-distributions are plotted along the corresponding axes (at the outside of the graphs). Colors separate the φ/ψ-combinations with respect to its populations, *i.e.*, the grey part contains 10% of all φ/ψ-combinations with the lowest populations, while yellow represents the 10% with the highest population, and the colors in between represent always 10% of all φ/ψ-combinations with corresponding populations.

bonds present in the crystal structure as well as in the native and ester-linked HEWL simulations. All calculations are based on the first 3.5 ns of the simulations. The H-bonds involving the peptide NH moieties as donors and those with an occurrence of less than 10% were not listed.

Overall, the degree of H-bonding involving side chains is not changing significantly upon introduction of ester linkages (see *Table 2*). But, the pattern of H-bonds does change. Only *ca.* 25% of the native H-bonds is also seen in the ester-linked simulation and *vice versa*. Thus, one source of the loss of stability is the absence of the 126 peptide NH H-bond donors in ester-linked HEWL. These are replaced by 126 O_α acceptor

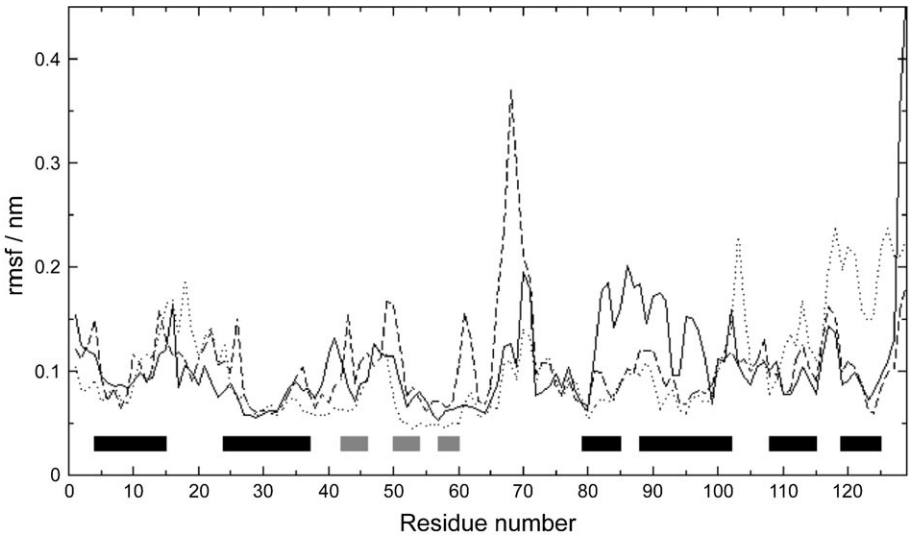

Fig. 5. *Root-mean-square positional fluctuations* (rmsf) *of* C_α*-atoms through a simulation time window of 2 ns* (solid line: ester-linked HEWL, 1.5–3.5 ns; dashed line: ester-linked HEWL, 8–10 ns; dotted line: native HEWL, 1.5–3.5 ns). The bars indicate the secondary structure of native HEWL observed in the crystal structure [15]: α-helix (black), β-bridge (thin, black), and β-strand (grey).

atoms, of which only two form a H-bond with side-chain atoms (Table S1 as *Supplementary Material*).

The result of the H-bond analysis depends on the geometric criterion used. However, the use of looser distance and angle bounds turns out not to change the overall picture of the differences in H-bonding between native and ester-linked HEWL.

Discussion. – *Table 1* and *Figs. 1–6* contain information on the evolution of the native and ester-linked protein structures with time, as expressed in the form of various quantities. The picture that emerges indicates that a simulation of 10 ns is more than sufficient to equilibrate ester-linked HEWL when starting from the X-ray structure of native HEWL. An equilibration period of 1–2 ns seems sufficient. The larger changes are expected for ester-linked HEWL, but its properties during the analysis periods 1.5–3.5 and 8–10 ns appear rather similar.

The observation that ester-linked HEWL does not unfold and seems to be more compact than native HEWL is in accordance with experimental studies on the distribution of end-to-end distances and loop formation of polypeptide chains [20–22]. The slight compaction might be due to the hydrophobic effect, which may be able to reduce the hydrophobic surface area when the intra-protein H-bonding is reduced due to the amide by ester-linkage replacement [10][11], and, second, it might be due to the changed network of H-bonds involving side chains. The former cause seems in contrast with the fact that the polypeptides studied by *Kiefhaber* and co-workers [17][19][20] are devoid of hydrophobic groups. But, compaction may have different causes for

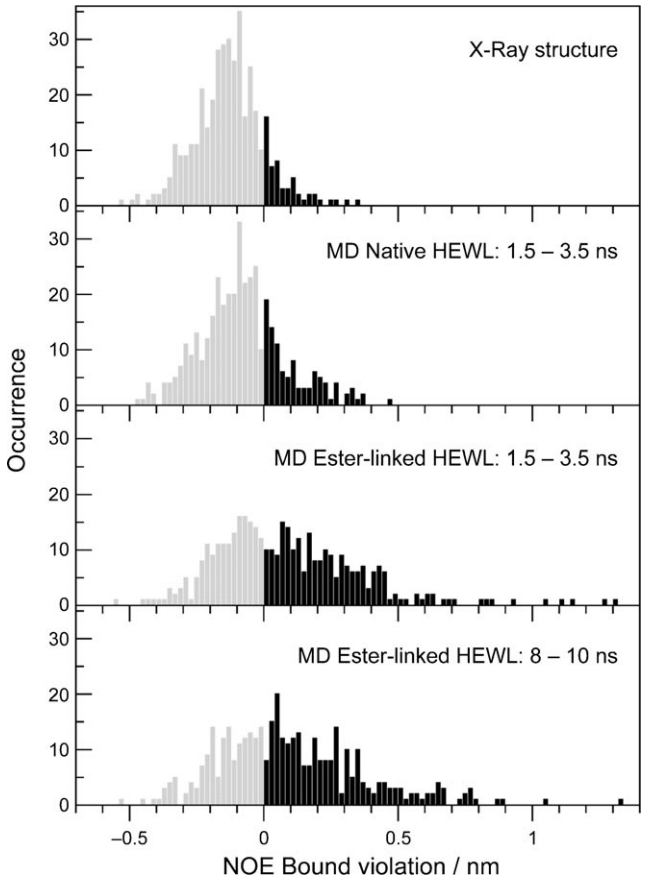

Fig. 6. *Occurrence of* r^{-3}*-averaged* $^1H-^1H$ *long-range NOE distances* (residues i to j with $j \geq i+4$) *for* (from top to bottom) *the X-ray structure* [15], *the native HEWL simulation* (1.5–3.5 ns), *and the ester-linked HEWL simulation* (1.5–3.5 and 8–10 ns). Positioning of the missing H-atoms of ester-linked HEWL were achieved according to the M procedure.

differently composed peptides. For this reason, we refrained from a detailed comparison with the results by *Kelly* and co-workers [10–13] for the PIN WW domain. The relative stability of the π-helix may be due to its smaller surface-to-volume ratio and its smaller H-bond-to-chain length ratio compared to an α-helix, which may make it less prone to destabilization due to reduced H-bonding.

We have considered performing a conformational clustering analysis of each MD trajectory. Because the results will heavily depend on the criterion used and are thus less instructive than an analysis of H-bonds, φ- and ψ-angles, and secondary structure, we have refrained from such an analysis.

Other perturbations than the replacement of amide linkages by ester linkages that would reduce the H-bonding capacity of the backbone have been considered. One could, for example, switch off all backbone H-to-O nonbonded interactions in the force

Table 1. *NOE Distance Bound Violations in the X-Ray Structure of Native HEWL and in the MD Simulation of Ester-Linked HEWL with Respect to the Experimental NMR NOE Distance Bounds* [18]. Two sets of NOE bounds have been used: set 1 includes 1630 NOE bounds, while set 2 contains only the 392 long-range (residues i and j with $j \geq i+4$) NOE bounds. In ester-linked HEWL, the virtual H-atom positions are generated using the M procedure. R_0 is the experimentally derived distance bound including pseudo-atom corrections [19]. R_E is the distance obtained from the MD trajectories using r^{-3} averaging.

Averaging period [ns] or structure	Number of violations (set 1/set 2):		
	> 0.1 nm	> 0.3 nm	> 0.5 nm
X-ray crystal	70/46	11/9	0/0
1.5 – 3.5 ns	241/162	83/72	22/21
8.0 – 10.0 ns	252/169	91/79	33/31
0.0 – 10.0 ns	205/150	46/41	12/12
List of the 12 largest NOE violations in ester-linked HEWL:			
NOE pair	Exp. R_0 [nm]	R_E [nm]	$R_E - R_0$ [nm]
$H_N(13Lys) - H_\delta(129Leu)$	0.69	1.27	0.58
$H_{\varepsilon_3}(28Trp) - H_\delta(88Ile)$	0.90	1.62	0.72
$H_\beta(95Ala) - H_\eta(108Trp)$	0.40	1.16	0.76
$H_\beta(95Ala) - H_{\zeta_2}(108Trp)$	0.45	1.17	0.72
$H_\alpha(95Ala) - H_{\zeta_2}(108Trp)$	0.30	1.12	0.82
$H_{\gamma_2}(98Ile) - H_N(107Ala)$	0.70	1.22	0.52
$H_\varepsilon(12Met) - H_\gamma(17Leu)$	0.30	0.89	0.59
$H_\varepsilon(12Met) - H_\alpha(28Trp)$	0.45	1.19	0.74
$H_\varepsilon(12Met) - H_{\varepsilon_3}(28Trp)$	0.45	1.04	0.59
$H_\varepsilon(12Met) - H_{\delta_1}(28Trp)$	0.45	1.08	0.63
$H_{\gamma_2}(98Ile) - H_\beta(107Ala)$	0.75	1.37	0.62
$H_N(98Ile) - H_{\zeta_2}(108Trp)$	0.55	1.11	0.56

Table 2. *Overview of the H-Bonds* (number and occurrence during the first 3.5 ns of simulation time) *Not Involving Peptide H-Atoms of the Backbone of HEWL.* The H-bonds with an occurrence lower than 10% are not considered. It is also shown how many H-bonds occur in both of the two MD simulations. The occurrence is the cumulative sum over the corresponding number of H-bonds, the full list is available as *Supplementary Material*.

System	Backbone/side chain		Side chains only	
	Number	Occurrence [%]	Number	Occurrence [%]
X-Ray	18	1800	15	1500
MD Native	26	686	26	1033
MD Ester-linked	22	499	34	1013
MD Native and ester-linked	5	109/184	9	468/324

field. However, this would convert the N–H with C–O dipole–dipole nonbonded interaction into a N with C–O charge–dipole interaction, which is a larger perturbation than the amide to ester (NH to O) one. For this reason, we have refrained from purely computational perturbations and only considered those that can be physically executed. These offer the possibility of falsification of our results. To this end, experimental studies like those of [10][11] could be performed for HEWL or observables measurable by NMR or other spectroscopic techniques could be

determined for ester-linked HEWL. In particular, it would be interesting to measure proton–proton NOEs for ester-linked HEWL and compare the observed pattern of atom–atom contacts to the one reported here.

Conclusions. – The role of the peptide NH groups in maintaining the three-dimensional structure of the protein HEWL has been investigated by replacing all but two (70 Pro and 79 Pro) peptide linkages of amino acid residues by ester linkages. The structures obtained by MD computer simulation were compared to those of a comparable, previously published MD simulation of the native protein and to the X-ray crystal structure of lysozyme.

The structure of the ester-linked protein rapidly moves away from that of the native one, but does not completely unfold. Its surface area and radius of gyration (rgyr) indicate a slight compaction compared to the native protein in solution. Due to the differences between peptide and ester linkages, the distributions of backbone φ- and ψ-angles are different. The extent of backbone–side chain and side chain–side chain H-bonding does not change upon replacement of peptide linkages by ester linkages, but the pattern of these H-bonds does change: only *ca.* 25% of the H-bonds in the native simulation is also observed in the ester-linked simulation and *vice versa*. The typical secondary structure of lysozyme, three β-strands, and four α-helices and two short 3_{10}-helices, are largely lost in the ester-linked lysozyme.

The atom-positional fluctuations do not significantly increase upon esterification of the peptide bonds, but differences are observed for particular residues.

The overall picture is that ester-linked lysozyme remains folded, due to the presence of four disulfide bridges, H-bonds involving side chains, and its hydrophobic core of side-chain moieties. Because it is experimentally feasible to synthesize ester-linked lysozyme, these observations from computer simulation eagerly await falsification by experiment.

Computational Methods

Molecular Model. The MD simulation of ester-linked HEWL (10 ns in H_2O) was performed using the GROMOS96 [23] simulation package and the GROMOS force-field parameter set 45A3 [24].

The initial structure was derived from a (native) crystal structure [15] by replacing the peptide bonds by ester bonds using force-field parameters as described by *Gattin et al.* [25]. This replacement was conducted for 126 of the 129 residues, all but 70 Pro, 79 Pro, and the terminal NH_3^+. The system was solvated in a periodic, cubic box (box lengths 8.54 nm) containing 19,675 SPC H_2O [26] molecules. The protonation states of protonizable amino acids corresponded to a pH of 7. This resulted in a system of in total 60,228 atoms including eight Cl^- ions to preserve the overall neutrality of the system.

Simulation Protocol. At the beginning of the simulation, the velocities of the atoms were assigned from a *Maxwell* distribution at 300 K. After a thermalization and equilibration (100 ps), the MD simulation was performed at constant temp. (300 K) using the weak coupling algorithm [27] with corresponding relaxation time $\tau = 0.1$. All bond lengths and the bond angles of the H_2O molecules were kept rigid by applying constraints using the SHAKE algorithm [28] with a relative geometric tolerance of 10^{-4}. This allowed an integration time step of 2 fs when solving the equations of motion of the system using the leap-frog algorithm [29]. Triple-range cutoff radii were used to treat long-range interactions (*Van der Waals* and electrostatic): interactions within the short-range cutoff (0.8 nm) were calculated every time step from a pair list that was generated every fifth step, when also interactions between 0.8 and 1.4 nm were computed. The long-range electrostatic forces were represented by a reaction field [30] with

a relative permittivity of $\varepsilon_{RF} = 61$ [31] outside the long-range cutoff of 1.4 nm. The centre of mass motion was removed every 2 ps. Configurations of the system were saved every 0.4 ps for analysis.

Analysis. All analyses were performed using the tools of the GROMOS96 simulation software [23]. They are based on two different MD simulations: the ester-linked HEWL simulation described above (10 ns) and an earlier performed (native) HEWL simulation (3.5 ns) [14] based on the same force-field parameter set (45A3) and initial structure [15].

A straightforward comparison of structures of the ester-linked protein with those of the native one is hampered by the absence of the peptide H-atoms in the ester-linked protein. Therefore, we considered three slightly different procedures to generate virtual H-atom positions using the positions of the ester O_α atom, and its covalently bound neighbors C and C_α in the ester linkage $-CO-O_\alpha-C_\alpha-$, where CO denotes the carbonyl group. All three procedures put the virtual H-atom such that the two angles $C-O_\alpha-H$ and $H-O_\alpha-C_\alpha$ have the same value, and the $O_\alpha-H$ distance is 0.1 nm. The three methods differ with respect to the plane in which the H-atom will be placed. The COC procedure places the virtual H-atom in the $C-O_\alpha-C_\alpha$ plane, attached to the O_α atom. The OCO procedure places the virtual H-atom in the $O-C-O_\alpha$ plane, and the so-called M procedure places the virtual H-atom in the plane bisecting the $C-O_\alpha-C_\alpha$ and $O-C-O_\alpha$ planes. If the dihedral angle ω ($C_\alpha-C-O_\alpha-C_\alpha$) is 180° (or 0°), all three procedures yield the same H-atom position. The dihedral angle distribution of the $\omega(C_\alpha-C-O_\alpha-C_\alpha)$ dihedral angle is shown in *Fig. 7*. It is much broader than the distribution of the $\omega(C_\alpha-C-N-C_\alpha)$ dihedral angle in native lysozyme. This means that the H-atom positions generated by the three procedures may show differences for the ω-values in the tails of the distribution. However, a comparison of the three procedures with respect to structural properties obtained showed that they yield comparable results. Therefore, we only use the M procedure to place the virtual H-atoms.

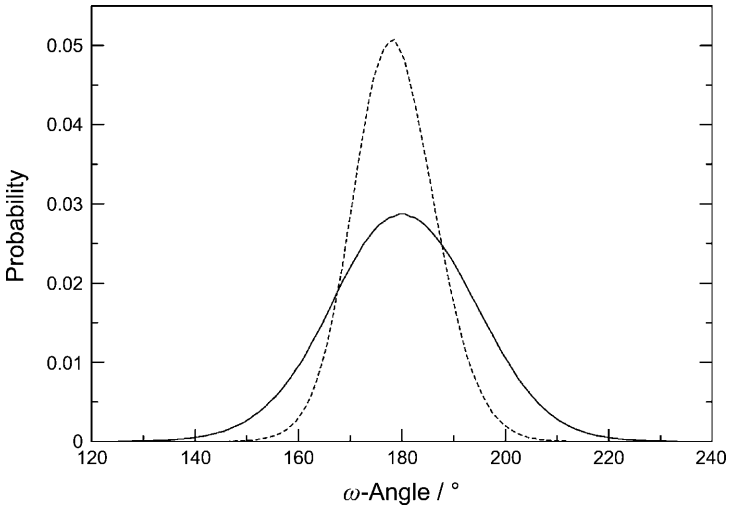

Fig. 7. *The ω angle* (native: $C_\alpha-C-N-C_\alpha$, dashed line; ester-linked: $C_\alpha-C-O_\alpha-C_\alpha$, solid line) *is calculated from the 3.5 and 10 ns MD simulation of native HEWL and ester-linked HEWL, respectively.* Calculations were performed for all residues except the two proline residues 70 and 79.

The atom-positional root-mean-square differences (rmsd) between pairs of structures have been evaluated based on all backbone atoms (native: N, C, O, C_α; ester-linked: O_α, C, O, C_α) of all 129 residues. The atom-positional rms fluctuations (rmsf) were also calculated for all 129 residues, but only considering the C_α atoms of the backbone. In both cases (rmsd and rmsf), a translational superposition of the centers of mass of the protein and a least-squares rotational fit with respect to the crystal structure was applied. The solvent-accessible surface area (sasa) was computed using the algorithm proposed by

Lee and *Richards* [32]. The volume of the protein was calculated in the following way: the structure of the protein (every time frame separately) was superposed on to an equilibrated box of pure SPC H_2O. All H_2O molecules overlapping with the atoms of the protein were removed from the box. Knowing the number of removed H_2O molecules and the average volume of one H_2O molecule, the protein volume can be estimated.

The radius of gyration of the protein (rgyr), a measure of the compactness of the structure that can be related to light-scattering intensity, was calculated using the definition

$$R_{\text{gyr}} = \left[\frac{1}{N_\alpha}\sum_{i=1}^{N_\alpha}(\vec{r}_i - \vec{r}_{\text{cm}})^2\right]^{1/2} \tag{1}$$

with

$$\vec{R}_{\text{cm}} = M^{-1}\sum_{i=1}^{N_\alpha} m_i\vec{r}_i \tag{2}$$

and

$$M = \sum_{i=1}^{N_\alpha} m_i, \tag{3}$$

in which $\vec{r}_i$ denotes the Cartesian position vector of atom i, m_i is its mass, and N_α denotes the number of protein atoms.

The H-bonds were analyzed according to a geometric criterion: a minimum donor – hydrogen – acceptor angle of 135° and a maximum hydrogen – acceptor distance of 0.25 nm [23]. Secondary-structure assignments were performed using the rules defined by *Kabsch* and *Sander* [33].

Proton – proton distances were used to compare the simulations to experimental NMR data for native HEWL. These comprise upper bounds for distances between pairs of atoms that can be derived from intensities of peaks observed in NMR nuclear *Overhauser* effect (NOE) spectra. Out of a set of 1632 NOE upper bounds [18], 1630 were compared to the calculated $^1H-^1H$ distances averaged using $1/r^3$ averaging, $\bar{r} = (\langle r^{-3}\rangle)^{-1/3}$. Interproton distances involving (aliphatic) H-atoms not treated explicitly in the simulations were calculated by defining virtual and pseudo-atoms [19][23] during the analysis as described in [14][17].

This work was financially supported by the *National Center of Competence in Research* (*NCCR*) *in Structural Biology* and by grant No. 200020-121913 of the *Swiss National Science Foundation*, and by grant number 228076 of the *European Research Council*, which is gratefully acknowledged.

REFERENCES

[1] L. S. Itzhaki, P. G. Wolynes, *Curr. Opin. Struct. Biol.* **2010**, *20*, 1.
[2] R. L. Baldwin, *J. Mol. Biol.* **2007**, *371*, 283.
[3] M. M. Schemjakin, *Angew. Chem.* **1960**, *72*, 342.
[4] M. M. Schemjakin, N. A. Aldanova, E. I. Vinogradova, M. Y. Feigina, *Tetrahedron Lett.* **1963**, *4*, 1921.
[5] R. L. Hamill, C. E. Higgens, H. E. Boaz, M. Gorman, *Tetrahedron Lett.* **1969**, *10*, 4255.
[6] K. Takesako, H. Kuroda, T. Inoue, F. Haruna, Y. Yoshikawa, I. Kato, K. Uchida, T. Hiratani, H. Yamaguchi, *J. Antibiot.* **1993**, *46*, 1414.
[7] E. A. Gallo, S. H. Gellman, *J. Am. Chem. Soc.* **1993**, *115*, 9774.
[8] E. M. Arnett, E. J. Mitchell, T. S. S. R. Murty, *J. Am. Chem. Soc.* **1974**, *96*, 3875.
[9] J. N. Spencer, R. C. Garrett, F. J. Mayer, J. E. Merkle, C. R. Powell, M. T. Tran, S. K. Berger, *Can. J. Chem.* **1980**, *58*, 1372.
[10] S. Deechongkit, H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, J. W. Kelly, *Nature* **2004**, *430*, 101.
[11] S. Deechongkit, P. E. Dawson, J. W. Kelly, *J. Am. Chem. Soc.* **2004**, *126*, 16762.

[12] S. Deechongkit, H. Nguyen, M. Jager, E. T. Powers, M. Gruebele, J. W. Kelly, *Curr. Opin. Struct. Biol.* **2006**, *16*, 94.
[13] M. Jager, S. Deechongkit, E. K. Koepf, H. Nguyen, J. Gao, E. T. Powers, M. Gruebele, J. W. Kelly, *Biopolymers* **2008**, *90*, 751.
[14] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, W. F. van Gunsteren, *J. Biomol. NMR* **2004**, *30*, 407.
[15] P. J. Artymiuk, C. C. F. Blake, D. W. Rice, K. S. Wilson, *Acta Crystallogr., Sect. B* **1982**, *38*, 778.
[16] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.
[17] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur. Biophys. J.* **2005**, *34*, 273.
[18] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, *Protein Sci.* **2001**, *10*, 677.
[19] K. Wüthrich, M. Billeter, W. Braun, *J. Mol. Biol.* **1983**, *169*, 949.
[20] A. Möglich, K. Joder, T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12394.
[21] B. Fierz, T. Kiefhaber, *J. Am. Chem. Soc.* **2007**, *129*, 672.
[22] B. Fierz, H. Satzger, C. Root, P. Gilch, W. Zinth, T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2163.
[23] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, 'Biomolecular Simulation: The GROMOS96 Manual and User Guide', vdf Hochschulverlag AG an der ETH Zürich, Zürich, and BIOMOS b.v., Groningen, 1996.
[24] L. D. Schuler, X. Daura, W. F. van Gunsteren, *J. Comput. Chem.* **2001**, *22*, 1205.
[25] Z. Gattin, A. Glättli, B. Jaun, W. F. van Gunsteren, *Biopolymers* **2007**, *85*, 318.
[26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, in 'Interaction models for water in relation to protein hydration', Reidel, Dordrecht, 1981, p. 331 – 342.
[27] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.
[28] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
[29] R. W. Hockney, J. W. Eastwood, 'Computer simulation using particles', McGraw-Hill, New York, 1981.
[30] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451.
[31] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2001**, *115*, 1125.
[32] B. Lee, F. M. Richards, *J. Mol. Biol.* **1971**, *55*, 379.
[33] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.

Received February 25, 2010