A Structural Ensemble of Hen Egg-White Lysozyme in Aqueous Solution Based on 2043 NMR NOE, ^3J-coupling and S^2 Order-Parameter Data

Lorna J. Smith[†], Wilfred F. van Gunsteren[#] and Niels Hansen[&*]

[†]Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QR, UK,

[#]Institute of Molecular Physical Sciences, Swiss Federal Institute of Technology, ETH, CH-8093 Zurich, Switzerland,

[&]Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, D-70569 Stuttgart, Germany

[*]Corresponding author. Electronic mail: hansen@itt.uni-stuttgart.de

(tel. +49 711 685 66112)

Version/Date: 17-11-2025

For submission to J. Phys. Chem. B

Keywords

Structure determination, nuclear magnetic resonance, configurational averaging, flexibility, restraining forces.

Abstract

A set of 1630 NOE atom – atom distance upper bounds, 213 ^{3}J-coupling constant values and 200 S^2 order-parameter values derived from experimental NMR data is used in molecular dynamics simulations of Hen Egg-White Lysozyme (HEWL) in aqueous solution in order to generate a Boltzmann-weighted structural ensemble for the protein in aqueous solution that is compatible with the 2043 NMR data. The two protein force fields used, the GROMOS 54A7 and 54A8 force fields, which only differ in the partial charges of the charged side chains and the chain termini of the protein, show comparable behaviour. Analysis of the generated structural ensembles shows that the protein in aqueous solution adopts a greater variety of hydrogen-bond patterns than is suggested by the various X-ray crystal structures of the protein.

1. Introduction

Structural information on proteins can be derived from a variety of observable quantities, such as X-ray or electron diffraction intensities, NMR, CD, Raman or infrared spectra to mention a few[1]. Although X-ray or electron diffraction of crystals has a high information density, that is, the ratio of the number of independent, measured values of observable quantities for a molecule and the number of independent molecular degrees of freedom, these techniques do not provide detailed information on the ensemble of configurations characterising the protein in solution. Of all techniques available to obtain information on proteins in solution, NMR shows the highest information density. In order to derive structural information from values of observable quantities Q measurable by NMR, such as Nuclear Overhauser effect (NOE) intensities, ^3J-coupling constants, chemical shifts, relaxation quantities such as S^2 order parameters, or residual dipolar couplings (RDCs), a relation $Q(\vec{r}^{\,N})$ between the observable quantity Q and the structure or configuration $\vec{r}^{\,N}$ of the protein is required. Here $\vec{r}^{\,N} \equiv (\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)$ denotes the $3N$ Cartesian coordinates of the N atoms of the protein or in the molecular system considered.

The measured value Q^{exp} of such an observable quantity Q is an average $\langle Q \rangle_{space,time}$ of Q over the molecules (space) in the test tube and over a time window determined by the particular measurement technique. This means that $\langle Q \rangle$ constitutes an average over a Boltzmann-weighted set, i.e. a statistical-mechanical ensemble, of configurations. The weights are proportional to $\exp(-V(\vec{r}^{\,N})/(k_B T))$, where $V(\vec{r}^{\,N})$ indicates the energy of a molecular configuration or structure $\vec{r}^{\,N}$, k_B is Boltzmann's

constant and T is the temperature.

If an observable quantity Q is dependent on a single molecular configuration $\vec{r}^{\,N}$, one may try to derive an expression or function $Q(\vec{r}^{\,N})$ that approximates the relation between Q and $\vec{r}^{\,N}$, which expression may then be used to derive molecular structures that are compatible with measured values of Q, i.e.

$$< Q >= \int Q(\vec{r}^{\,N}) \exp(-V(\vec{r}^{\,N}) / (k_B T)) d\vec{r}^{\,N} \,/\, \int \exp(-V(\vec{r}^{\,N}) / (k_B T)) d\vec{r}^{\,N}, \qquad (1)$$

i.e. Boltzmann-weighted averages over the, $3N$-dimensional molecular configuration space.

The intensity $I(r_{k_1 k_2})$ of a NOESY or ROESY signal is related to the distance between the two atoms k_1 and k_2 involved in the NOE and can be used to define NOE distance upper bounds[2,3]. A ^3J-coupling $^3J_{k_1 k_4}$ involving atoms k_1 and k_4 can be related to a torsional angle $\varphi_{k_1 k_2 k_3 k_4}$ in the protein using the Karplus relation[4,5]. The function $^3J_{k_1 k_4}(\varphi(\vec{r}_{k_1}, \vec{r}_{k_2}, \vec{r}_{k_3}, \vec{r}_{k_4}))$ depends on empirically determined parameters, see for example[1,6]. The relation between a chemical shift and structure $\sigma_k(\vec{r}^{\,N})$ depends on more than local covalent structure around atom k and is based on folded as well as random structures. It cannot reliably be formulated as a simple mathematical function of a few atom positions that rests on a sound physical (not empirical-statistical) basis, see for example[1,7,8]. S^2 order parameters contain dynamical information and are defined in terms of local, small oscillations of particular vectors or bonds in the protein[9-11]. RDCs are the result of averaging a second-order Legendre function of the angle θ between the magnetic field and a particular vector or bond in the protein over the complete rotation of the protein immersed in a medium that induces a slightly non-

uniform or anisotropic rotational distribution. Since this rotational distribution cannot be experimentally measured, an RDC-value is commonly calculated by minimising the difference between calculated and measured, target, RDC values. This procedure leads to a dependence of calculated RDC values upon assumptions about the shape of the immeasurable rotational distribution and on the size, type and distribution over the molecule of the set of measured RDCs used as target RDC values. It introduces uncertainty into the calculated RDC values and renders RDCs less useful for structure determination of proteins[12,13]. For this reason RDC values will not be applied as restraints in the current study.

Structure determination of proteins requires the definition of a molecular model, because only in case of X-ray diffraction of small molecules there may be sufficient data, i.e. diffraction intensities, to directly determine the atom positions without any assumptions about bond lengths, bond-angles, etc. This implies that the lower the information density of a measurement, the more the generated structure will depend on the quality of the molecular model used. Thus, when determining structure based on NMR data, the molecular model or force field used will significantly contribute to the quality of the obtained structure[12].

In the current article, Boltzmann-weighted ensembles of structures or configurations of hen egg-white lysozyme (HEWL) in aqueous solution are generated using molecular dynamics (MD) simulation in aqueous solution with application of time-averaged restraints, 1630 NOE atom – atom distance upper bound, 213 ^{3}J-coupling and 200 S^2 order-parameter values, and, also without any restraining. In

addition, two different force fields, the GROMOS force fields 54A7[14,15] and 54A8[16] are applied.

HEWL (see Figure 1) is one of the most studied and best characterised proteins. It has been used as a model system for a wide range of studies including investigations of protein folding, aggregation, amyloid formation, enzyme mechanisms and nanoparticle interactions.[17-21] HEWL was the first enzyme to have its structure determined by X-ray diffraction[22] and now there are over 100 independently determined X-ray structures of HEWL in the crystalline environment available. However, most of the studies and applications of HEWL use the protein in solution. NMR solution structures of the protein have been determined[23,24] but at the time of the structure determinations, refinement of the structure in aqueous solution was not usually done as the final step of an NMR structure determination. In the past few years we have extensively studied and developed approaches for using experimental NMR data as restraints in MD structure refinement using HEWL as a model system.[25-27] Here we bring together this work in a final comprehensive refinement to give a structural ensemble of HEWL in aqueous solution. This ensemble enables us to address various questions about the structure of the protein in solution raised during the experimental NMR structure determination,[24] particularly the nature of the 3_{10} helices and the behaviour of the more mobile loop and turn regions in solution.

2. Atom-atom distance, ^3J-coupling and S^2 order-parameter restraints in molecular simulation

2.1. Choice of restraining function

A simple general restraining function for a quantity Q_k, continuous with continuous derivative, is a quadratic one with a flat bottom of size $2\Delta Q^{fb}$ that allows for a penalty-free range of deviations $\pm\Delta Q^{fb}$ of the average $\langle Q_k \rangle$-value from its target (measured) value Q_k^0. In an MD simulation the time average, $\overline{Q_k}^t \equiv \langle Q_k \rangle_t$, is used. For large deviations of $\overline{Q_k}^t$ from Q_k^0, the restraining function is chosen to be linear in order to avoid large restraining forces and energies. Such a restraining function[1] would read for $\overline{Q_k}^t > Q_k^0$,

$$V_k^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0,K^Q,\Delta Q^{fb},\Delta Q^h\right)=\frac{1}{2}K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0-\Delta Q^{fb}\right)^2$$
$$H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0+\Delta Q^{fb}\right)\left(1-H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0+\Delta Q^{fb}+\Delta Q^h\right)\right)$$
$$+K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0-\Delta Q^{fb}-\frac{1}{2}\Delta Q^h\right)\Delta Q^h H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0+\Delta Q^{fb}+\Delta Q^h\right), \quad (2a)$$

and for $\overline{Q_k}^t < Q_k^0$,

$$V_k^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0,K^Q,\Delta Q^{fb},\Delta Q^h\right)=\frac{1}{2}K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0+\Delta Q^{fb}\right)^2$$
$$\left(1-H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0-\Delta Q^{fb}\right)\right)H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0-\Delta Q^{fb}-\Delta Q^h\right)$$
$$-K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0+\Delta Q^{fb}+\frac{1}{2}\Delta Q^h\right)\Delta Q^h\left(1-H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0-\Delta Q^{fb}-\Delta Q^h\right)\right), \quad (2b)$$

with the Heaviside step function $H\left(x;x_0\right)$ defined by

$$H\left(x;x_0\right)=0 \quad \text{for } x < x_0, \tag{3a}$$

$$H\left(x;x_0\right)=1 \quad \text{for } x \geq x_0. \tag{3b}$$

The corresponding force on $\overline{Q_k}^t$, that is, the negative of the derivative of $V_k^{Q,restr}$ with respect to $\overline{Q_k}^t$ is then for $\overline{Q_k}^t > Q_k^0$,

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=0 \qquad \text{for } \overline{Q_k}^t < Q_k^0+\Delta Q^{fb}, \tag{4a}$$

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=-K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0-\Delta Q^{fb}\right)$$

$$\text{for } Q_k^0+\Delta Q^{fb} \leq \overline{Q_k}^t \leq Q_k^0+\Delta Q^{fb}+\Delta Q^h, \tag{4b}$$

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=-K^Q\Delta Q^h \qquad \text{for } \overline{Q_k}^t > Q_k^0+\Delta Q^{fb}+\Delta Q^h, \tag{4c}$$

and for $\overline{Q_k}^t < Q_k^0$,

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=0 \qquad \text{for } \overline{Q_k}^t < Q_k^0-\Delta Q^{fb}, \tag{4d}$$

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=-K^Q\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0+\Delta Q^{fb}\right)$$

$$\text{for } Q_k^0-\Delta Q^{fb}-\Delta Q^h \leq \overline{Q_k}^t \leq Q_k^0-\Delta Q^{fb}, \tag{4e}$$

$$f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)=+K^Q\Delta Q^h \qquad \text{for } \overline{Q_k}^t < Q_k^0-\Delta Q^{fb}-\Delta Q^h. \tag{4f}$$

To obtain the force on a particle *i*, these expressions are multiplied by $\partial\overline{Q_k(r^N(t))}^t \Big/ \partial\vec{r}_i(t)$, the derivative of $\overline{Q_k}^t$ with respect to $\vec{r}_i(t)$.

For a set of $k = 1, 2, \ldots, N_Q$ Q_k-values we have

$$V^{Q,restr}\left(\overline{Q\left(\vec{r}^{\,N}\right)}^t\right)=\sum_{k=1}^{N_Q}V_k^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0,K^Q,\Delta Q^{fb},\Delta Q^h\right). \tag{5}$$

Here, K^Q is the restraining force constant.

The GROMOS force fields[14-16,28] treat aliphatic CH_n moieties as united atoms.

If explicit aliphatic hydrogen atom positions are required, the implementation allows for the use of a virtual H atom[28-30].

2.2. Accounting for time-averaging

When restraining an averaged quantity in a MD simulation, the averaging cannot be over the whole simulation period *t*, because the contribution of the configuration at time *t* to the average would tend to zero for *t* approaching infinity, which implies that the restraining force, which can only be applied to the current configuration at time *t*, would also tend to zero for *t* approaching infinity. For this reason, instead of using the time average $\overline{Q_k}^t \equiv \langle Q_k \rangle_t$, an exponential memory function is used in the time-average, $\overline{Q_k}^{t,\exp}$, that is used in the restraining function[31,32],

$$\overline{Q_k\left(\vec{r}^{\,N}(t)\right)}^t = \left(\tau_Q\left(1-\exp(-t/\tau_Q)\right)\right)^{-1}\int_0^t \exp(-(t-t')/\tau_Q)Q_k(\vec{r}^{\,N}(t'))dt', \qquad (6)$$

where τ_Q is the memory relaxation time, which determines the averaging time period.

2.3. Calculation of the restraining force

The restraining force on atom *i* with position vector $\vec{r}_i$ is, using $f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)$ from Eq. (4),

$$\vec{f}_i^{\,Q,restr}(t) = -\frac{\partial V^{Q,restr}(t)}{\partial \vec{r}_i(t)} = \sum_{k=1}^{N_Q} f^{Q,restr}\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t\right)\frac{\partial \overline{Q_k(r^N(t))}^t}{\partial \vec{r}_i(t)}. \qquad (7)$$

In a simulation, the atomic configurations are separated by a time step or interval Δt, so in discretised form we have for the n^{th} time step for the exponentially damped time-averaged quantity in Eq. (6)

$$\overline{Q_k\left(n\Delta t\right)}^{n\Delta t,\exp} = Q_k\left(n\Delta t\right)\left(1-\exp(-\Delta t/\tau_Q)\right)+\exp(-\Delta t/\tau_Q)\overline{Q_k\left((n-1)\Delta t\right)}^{(n-1)\Delta t,\exp} \quad , \quad (8)$$

and using this equation

$$\frac{\partial\overline{Q_k(n\Delta t)}^t}{\partial Q_k(n\Delta t)} = 1-\exp(-\Delta t/\tau_Q) \equiv X \; . \quad (9)$$

Generally, the memory relaxation time τ_Q is chosen to be much longer than the MD integration time step Δt,

$$\Delta t \ll \tau_Q \ll t^{MD} \quad , \quad (10)$$

where t^{MD} is the length of the MD simulation. This means that the derivative Eq. (9) can be approximated by $\Delta t/\tau_Q$. However, when restraining NOE atom-atom distances and S^2 order parameters, the derivative in Eq. (9), $X \equiv 1-\exp(-\Delta t/\tau_Q)$, is set to 1[11,33]. Use of Eq. (9) would mean that the restraining force is inversely proportional to the exponential memory relaxation time τ_Q (for $\Delta t \ll \tau_Q$). In order to obtain comparable restraining forces for different values of τ_Q, one would have to choose the force constant K^Q proportional to $X^{-1} \equiv (1-\exp(-\Delta t/\tau_Q))^{-1}$. Since the choice of the value for the force constant K^Q is empirical anyway, the factor X may be set to 1.

2.4. Calculation of *Q*-values using configurational averaging

The calculation of *Q*-values $\overline{Q_k(t)}^t$ is straightforward, averaging over MD trajectory structures. In the restraining simulations the *Q*-restraining forces on the molecule are calculated at every time step using an exponential damping factor in the average, using $\overline{Q_k(t)}^{t,\exp}$ for $\overline{Q_k(t)}^t$. These two types of averaged *Q*-values, calculated from $\overline{Q_k(t)}^{t,\exp}$ on the one hand and from $\overline{Q_k(t)}^t$ on the other, will differ. In order to analyse all trajectories in the same way, $\overline{Q_k(t)}^t$ is used when reporting Q-values calculated from the trajectories.

2.5. NOE atom-atom distance restraining

When using NOE cross-peak intensities for atom-atom distance restraining, generally only distance upper bounds are used, Eqs. (2a) and (4a-c). Lower bounds are not needed, the repulsive non-bonded interaction already provides a lower bound to atom-atom distances. Time-averaged $< r_{k_1 k_2}^{-3} >^{-1/3}$ NOE distance upper-bound restraints without a flat-bottom part (ΔQ^{fb} = 0 nm) are used in the distance-restraining function. In the present application to HEWL, $\Delta Q^{h} = \infty$ nm, $K^{NOE} = K^{Q}$ = 6000 kJmol^{-1}nm^{-2}, and the memory relaxation time $\tau_Q = \tau_{NOE}$ = 1 ns, much shorter than the rotational correlation time of 5.7 ns of HEWL in solution[23]. Because the GROMOS force fields make use of united atoms, positions of aliphatic hydrogen atoms were constructed based on standard geometries[1,28-30]. If an NOE upper bound involved non-stereo-specifically assigned protons, a pseudo atom was constructed[1,28,30,34]. The list of atom-atom distance upper bounds can be found in Supporting Information Table S1.

2.6. ^3J-coupling restraining

When applied to ^3J-coupling restraining, the use of the time-average $\overline{Q_k(t)}^t$ in the quadratic part of the restraining function $V_k^{Q,restr}$ (Eq. (2)) led to large structural fluctuations[35]. This is due to the fact that the average $\overline{Q_k(t)}^t$ of Q_k may lag behind the instantaneous value of $Q_k(t)$ of Q_k. If the restraining force only depends on $\overline{Q_k(t)}^t$, it may drive $Q_k(t)$ away from Q_k^0 in case $\overline{Q_k(t)}^t$ and $Q_k(t)$ are at different sides of Q_k^0. This effect is not observed in time-averaging NOE atom-atom distance restraining because (*i*) only attractive distance restraints are used, (*ii*) the r^{-3} or r^{-6} distance dependence of the NOE signal gives most weight to short distances when averaging,

thereby reducing the difference between $Q_k(t)$ and $\overline{Q_k(t)}^t$ for short distances $Q_k(t)$, and (*iii*) the repulsive non-bonded interaction between atoms prevents very small atom-atom distances $Q_k(t)$ being sampled. The problem of $\overline{Q_k(t)}^t$ lagging behind $Q_k(t)$ can be solved by using a biquadratic restraining function that depends on both $Q_k(t)$ and $\overline{Q_k(t)}^t$ [1,33],

$$V_k^{Q,restr,att}\left(Q_k(t),\overline{Q_k}^t;Q_k^0,K^Q,\Delta Q^{fb}\right)=\frac{1}{2}K^Q\left(Q_k(t)-Q_k^0-\Delta Q^{fb}\right)^2$$

$$H\left(Q_k(t);Q_k^0+\Delta Q^{fb}\right)\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0-\Delta Q^{fb}\right)^2 H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0+\Delta Q^{fb}\right), \quad (11a)$$

for $Q_k(t) > Q_k^0$ and $\overline{Q_k(t)}^t > Q_k^0$, and

$$V_k^{Q,restr,rep}\left(Q_k(t),\overline{Q_k}^t;Q_k^0,K^Q,\Delta Q^{fb}\right)=\frac{1}{2}K^Q\left(Q_k(t)-Q_k^0+\Delta Q^{fb}\right)^2$$

$$\left[1-H\left(Q_k(t);Q_k^0-\Delta Q^{fb}\right)\right]\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t-Q_k^0+\Delta Q^{fb}\right)^2\left[1-H\left(\overline{Q_k\left(\vec{r}^{\,N}\right)}^t;Q_k^0-\Delta Q^{fb}\right)\right], \quad (11b)$$

for $Q_k(t) < Q_k^0$ and $\overline{Q_k(t)}^t < Q_k^0$. Using this biquadratic form, the restraining function only generates a force when both the instantaneous value $Q_k(t)$ of Q_k and the time-averaged value $\overline{Q_k(t)}^t$ of Q_k lie outside the flat bottom of the restraining function $\overline{Q_k(t)}^t$.

The Karplus relation[4,5] expressing a ^3J-coupling in terms of the corresponding angle φ reads

$$^3J_{k_1k_4}(\varphi_{k_1k_2k_3k_4})=a\cos^2(\varphi_{k_1k_2k_3k_4})+b\cos(\varphi_{k_1k_2k_3k_4})+c \quad . \quad (12)$$

Below, the simplified notations $k \equiv (k_1,k_4)$ and $k \equiv (k_1,k_2,k_3,k_4)$ are used. The inverse $\varphi_k(^3J_k)$ of the Karplus relation, which expresses the angle φ_k as function of the corresponding ^3J_k-coupling, is multiple-valued. There are up to four values of the angle φ_k that may correspond to a single ^3J_k-coupling value. Depending on the initial value of

the angle φ_k, the restraining function $V_k^{Q,restr}\left(\overline{Q_k\left(\varphi_k\left(\vec{r}^{\,N}(t)\right)\right)}^{t}, Q_k\left(\varphi_k\left(\vec{r}^{\,N}(t)\right)\right); Q_k^0, K^Q, \Delta Q^{fb}, \Delta Q^h\right)$ may drive the angle φ_k in different directions to one of the different angle values that correspond to Q_k^0. If the target Q_k^0-value lies beyond one maximum or minimum of the function $Q_k(\varphi_k)$, but not beyond another maximum or minimum, the restraining function $V_k^{Q,restr}$ may drive the angle φ_k in the wrong direction. This problem can be solved by driving the structure out of the range of φ-values, for which $V_k^{Q,restr}$ has a low energy but where these energies cannot become zero.

A technique that achieves this, local-elevation sampling, has been proposed more than three decades ago[36]. It adds a memory function $V_k^{le}\left(\varphi_k(t)\right)$ depending on the torsional angle φ_k to the interaction or potential-energy function[25]. This memory function can have the form of a set of Gaussian functions $V_{k,i}^{le}\left(\varphi_k(t); \varphi_i^0, \Delta\phi^0\right)$ of width $\Delta\varphi^0$ centred at values φ_i^0 with i = 1, 2,.., N_{le}, for a given torsional angle φ_k. Thus,

$$V_k^{le}\left(\varphi_k(t)\right) = \sum_{i=1}^{N_{le}} V_{k,i}^{le}\left(\varphi_k(t); \varphi_i^0, \Delta\phi^0\right) \tag{13}$$

is a sum of N_{le} Gaussians centred at generally equidistant values φ_i^0, which are the same for each torsional angle k, with i=1, 2,.., N_{le} along the 360° range of φ_k values,

$$V_{k,i}^{le}\left(\varphi_k(t); \varphi_i^0, \Delta\phi^0\right) = K^{le}\omega_{k,i}(\varphi_k(t) : \varphi_i^0)\exp\left(-\frac{(\varphi_k(t)-\varphi_i^0)^2}{2(\Delta\varphi^0)^2}\right). \tag{14}$$

The local-elevation parameter N_{le} defines the number of intervals N_{le} of the torsional angle φ_k around the φ_i^0-values

$$\varphi_i^0 \equiv 2\pi i / N_{le}, \quad i = 1, 2, \ldots, N_{le} \tag{15}$$

and their widths

$$\Delta\varphi^0 \equiv 2\pi / N_{le} \quad . \tag{16}$$

K^{le} is the overall weight of the local-elevation contribution to the potential energy, while the weight factor $\omega_{k,i}(\varphi_k(t);\varphi_i^0)$ is

$$\omega_{k,i}(\varphi_k(t);\varphi_i^0) = (t)^{-1}\int_0^t \delta_{\phi_k(t')\varphi_i^0} \left\{ \begin{array}{l} V_k^{Q,restr,att}\left(Q_k(t'),\overline{Q_k}^{t'};Q_k^0,K^Q,\Delta Q^{fb}\right) \\ +V_k^{Q,restr,rep}\left(Q_k(t'),\overline{Q_k}^{t'};Q_k^0,K^Q,\Delta Q^{fb}\right) \end{array} \right\} dt' \tag{17}$$

with

$$\delta_{\varphi_k(t')\varphi_i^0} = 1 \;\text{ if }\; \varphi_i^0 - \Delta\varphi^0/2 \le \varphi_k(t') \le \varphi_i^0 + \Delta\varphi^0/2 \tag{18a}$$

and

$$\delta_{\varphi_k(t')\varphi_i^0} = 1 \;\text{ otherwise .} \tag{18b}$$

In this way the weight $\omega_{k,i}(\varphi_k(t);\varphi_i^0)$ is only increased when both $Q_k(t)$ and $\overline{Q_k(t)}^t$ lie outside the flat bottom of $V_k^{Q,restr}$.

When applying biquadratic local-elevation ^3J-coupling restraining to HEWL[25,26], ΔQ^{fb} = 1 Hz and $\Delta Q^h = \infty$ Hz, and N_{le} = 36, $K^Q = K^J$ = 50 kJmol^{-1}Hz^{-4} , and the memory relaxation time $\tau_Q = \tau_J$ = 500 ps. The lists of experimentally derived ^3J-coupling values can be found in Supporting Information Tables S2 – S5.

2.7. S^2 order-parameter restraining

S^2 order parameters can be calculated using the ensemble averaging expression[10,11]

$$S^2_{k_1k_2} = \frac{1}{2}\left\{ 3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\left\langle \frac{\mu_{k_1k_2\alpha}(t)\mu_{k_1k_2\beta}(t)}{r^3_{k_1k_2}(t)} \right\rangle^2_{t^{av}} - \left\langle \frac{1}{r^3_{k_1k_2}(t)} \right\rangle^2_{t^{av}} \right\} \left(r^{eff}_{k_1k_2}\right)^6 \quad , \tag{19}$$

where t^{av} indicates the time-averaging window, here 1 ns, shorter than the rotational correlation time of 5.7 ns of HEWL in solution[23],

$$\mu_{k_1k_2 1} \equiv (x_{k_1} - x_{k_2}) / r_{k_1k_2}\ ,\ \ \mu_{k_1k_2 2} \equiv (y_{k_1} - y_{k_2}) / r_{k_1k_2}\ ,\ \ \mu_{k_1k_2 3} \equiv (z_{k_1} - z_{k_2}) / r_{k_1k_2} \qquad (20)$$

are the x-, y-, and z-components of the vector $\vec{r}_{k_1k_2}$ and $r_{k_1k_2} \equiv \left|\vec{r}_{k_1k_2}\right|$ its length.[11] To obtain a dimensionless quantity the expression in curly brackets is multiplied with the 6th power of the effective length ($\vec{r}_{k_1k_2}$) of the vector $\vec{r}_{k_1k_2}$. Because in the present work bond length constraints are used, the length of $\vec{r}_{k_1k_2}$ is essentially constant over time and thus equal to its effective value $\vec{r}_{k_1k_2}$.

Before calculating $S^2_{k_1k_2}$, the HEWL protein trajectory structures are superimposed using the backbone atoms (N, C_α, C) of residues 3 – 126 in the fit in order to eliminate the effect of overall rotation of the protein upon the $S^2_{k_1k_2}$-values. Use of only the backbone atoms of four α-helices and two β-strands in HEWL did not lead to significantly different $S^2_{k_1k_2}$-values.

In the S^2 order-parameter restraining simulations, the restraining force constant $K^Q = K^S = 300$ kJmol^{-1}, $\Delta Q^h = \infty$ Hz and the flat-bottom parameter of the restraining potential-energy term is $\Delta Q^{fb} = \Delta S^2 = 0.1$, which means a flat bottom of 0.2 width[27]. The S^2 order parameter is calculated at every time step (2 fs) using an exponential damping factor in the average[11] with a memory relaxation time $\tau_Q = \tau_S = 200$ ps[27], and no rotational fit of the protein structures is carried out, which means that the calculated order parameters in the restrained simulation are influenced by the stochastic tumbling of the protein. These S^2 order-parameter values will thus differ slightly from the ones calculated from the saved trajectory structures, because in the averages in Eq. (19) trajectory structures 2 ps apart are used, the exponential damping factor is not used, the averaging period is $t^{av} = 1$ ns and a rotational fit of the protein structures is carried

out. However, to analyse all trajectories in the same way Eq. (19) was used for both the unrestrained and restrained trajectories. The lists of experimentally derived S^2 values can be found in Supporting Information Tables S6 – S9.

3. Molecular model, simulation set-up and analysis

The simulations were performed using the GROMOS simulation software package[28]. When treating realistic molecular systems, the use of Standard International (SI) units is recommended. Apart from restrictions when storing or printing data in non-exponential format, the GROMOS programs are independent of the chosen units. The units are defined by the ones used for the physical constants and atomic and molecular quantities to be specified in the PHYSICALCONSTANTS block in the GROMOS data files[28]. It is recommended that the following basic units are used: nanometer (nm) for length, atomic mass unit (u) for mass, picosecond (ps) for time, Kelvin (K) for temperature, and electronic charge (e) for charge. These basic units then determine the units of other quantities, e.g. kJ/mol for energy, kJ/(mol nm) for force, kJ/(mol nm^3) for pressure, and THz for frequency.

3.1. Molecular model and force field used for the protein

When solvated in explicit water, the protein was modelled using the GROMOS bio-molecular force fields 54A7[14,15] and 54A8[16]. The 54A8 force field involved a recalibration of the non-bonded interaction parameters for charged amino-acid side chains, based on ionic side-chain analogues. In view of the pH used in the experimental NMR measurements[23,37-39], pH=3.5[37-39] or 3.8[23], only Glu 35 was protonated and His was doubly protonated[40]. This means that the N-terminus was protonated (sum of partial charges +1) as well as the residues Lys 1, 13, 33, 96, 97, 116, Arg 5, 14, 21, 45, 61, 68, 73, 112, 114, 125, 128, and His 15 and Glu 35 (sum of partial charges 0). The C-terminus was deprotonated (sum of partial charges -1) as well as Glu 7, and Asp 18,

48, 52, 66, 87, 101, 119. This led to an overall charge of +10*e* (*e*: electronic charge) of the protein. In order to obtain an overall neutral system, 10 Cl^- ions were included in the solution. The simple point charge (SPC) model[41] was used to describe the solvent molecules in the rectangular periodic box. All bond lengths and the bond angle of the water molecules were kept rigid with a relative geometric precision of 10^{-4} using the SHAKE algorithm[42], allowing for a 2 fs MD time step in the leap-frog algorithm[43] used to integrate the equations of motion. For the non-bonded interactions a triple-range method[44] with cut-off radii of 0.8/1.4 nm was used. Short-range (within 0.8 nm) van der Waals and electrostatic interactions were evaluated every time step based on a charge-group pair list[28]. Medium-range van der Waals and electrostatic interactions, between pairs at a distance larger than 0.8 nm and shorter than 1.4 nm, were evaluated every fifth time step (10 fs), at which time point the pair list was updated, and kept constant between updates. Outside the larger cut-off radius (1.4 nm) a reaction-field approximation[45,46] with a relative dielectric permittivity of 61[47] was used. Minimum-image periodic boundary conditions were applied.

3.2. Simulation set-up

The X-ray crystal structure "2VB1" of the Protein Data Bank (PDB)[48], derived from a triclinic unit cell at 0.065 nm resolution at $T = 100$ K, was used as initial structure for the energy minimisations followed by MD simulations. It contains multiple side-chain conformations for 46 residues. For the initial structure the side-chain conformation with the highest occupancy was chosen.

The initial structure was first energy minimised in vacuo to release possible strain

induced by small differences in bond lengths, bond angles, improper dihedral angles, and short non-bonded contacts between the force-field parameters and the X-ray structure. Subsequently, the protein was put into a rectangular box filled with water molecules, such that the minimum solute-wall distance was 1.0 nm, and water molecules closer than 0.23 nm from the solute were removed. This resulted in a box with 12157 water molecules, of which the 10 water molecules with the highest electrostatic potential were replaced by Cl^- ions. Thus the initial protein structure was solvated by 12147 water molecules and 10 Cl^- ions. In order to relax unfavourable contacts between atoms of the solute and the solvent, a second energy minimisation was performed for the protein in the periodic box with water while keeping the atoms of the solute harmonically position-restrained[28] with a force constant of 25000 $kJmol^{-1}nm^{-2}$.

The resulting protein-water configuration was used as initial configuration for the MD simulations. In order to avoid artificial deformations in the protein structure due to relatively high-energy atomic interactions still present in the system, the MD simulations were started at T = 60 K and then the temperature was slowly raised to T = 308 K. Initial atomic velocities were sampled from a Maxwell distribution at T = 60 K. The equilibration scheme consisted of five short 20 ps simulations at temperatures 60, 120, 180, 240 and 308 K at constant volume. During the first four of the equilibration periods, the solute atoms were harmonically restrained to their positions in the initial structures with force constants of 25000, 2500, 250, and 25 $kJmol^{-1}nm^{-2}$. The temperature was kept constant using the weak-coupling algorithm[49] with a relaxation

or coupling time τ_T = 0.1 ps. Solute and solvent were separately coupled to the heat bath. Following this equilibration procedure, the simulations were performed at a reference temperature of 308 K and a reference pressure of 1 atm. The pressure was kept constant using the weak-coupling algorithm[49] with a coupling time τ_p = 0.5 ps and an isothermal compressibility κ_T = 4.575 10^{-4} $(kJmol^{-1}nm^{-3})^{-1}$. The centre of mass motion of the system was removed every 1000 time steps (2 ps).

3.3. NOE atom-atom distance restraints for HEWL

The GROMOS force fields treat aliphatic carbons as united CH, CH_2 and CH_3 atoms. So when calculating NOE distances, inter-hydrogen distances involving the aliphatic hydrogen atoms were calculated using virtual atomic positions for CH and pro-chiral CH_2[28-30] and pseudo-atomic positions for CH_3[34] for those hydrogen atoms. The pseudo-atom NOE distance bound corrections of Ref.[34] were used. The set of 1630 NOE atom – atom distance upper bounds for HEWL[23,24] can be found in Table S1 of Supporting Information. The NOE between Trp 28 HZ3 and Leu 56 HG was reassigned as between Trp 28 HZ3 and Leu 56 HD* following reassessment of the experimental spectra. Inter-hydrogen distances were calculated as $\langle r^{-3}\rangle^{-1/3}$, i.e. using r^{-3} averaging over the trajectory structures, where r indicates the actual hydrogen-hydrogen distance. The inter-hydrogen distances obtained from the simulations can be found in Table S1 of Supporting Information.

In view of the uncertainty due to the assumptions and approximations involved in the conversion of NOE signals to NOE atom – atom distance upper bounds for r^{-3} averaged distances, deviations from experiment of less than 0.1 nm are considered

insignificant.

3.4. ^3J-coupling restraints for HEWL

Two sets of experimentally derived backbone $^3J_{HN\text{-}H\alpha}$ couplings and two sets of experimentally derived side-chain $^3J_{H\alpha\text{-}H\beta}$ couplings of HEWL[25] were used, see Supporting Information Tables S2- S5.

1. A set (*bb1*) of 95 backbone $^3J_{HN\text{-}H\alpha}$-coupling values, see Table II of Ref.[37] from which the values for 11 glycine residues were omitted, because these had not been stereo-specifically assigned.
2. A set (*bb2*) of 22 experimentally stereo-specifically unassigned backbone $^3J_{HN\text{-}H\alpha}$-coupling values for the 11 glycine residues, see Table II of Ref.[37]. 10 of these were stereo-specifically assigned[26] based on a comparison of the $^3J_{HN\text{-}H\alpha}$-coupling values calculated from MD simulations and from X-ray structures.
3. A set (*sc1*) of 58 $^3J_{H\alpha\text{-}H\beta}$-coupling values, see Tables III and IV of Ref.[37], which were stereo-specifically assigned using experimental data.
4. A set (*sc2*) of 38 out of 40 experimentally stereo-specifically unassigned $^3J_{H\alpha\text{-}H\beta}$-coupling values, see Table III of Ref.[37], which were stereo-specifically assigned[26] based on the $^3J_{H\alpha\text{-}H\beta}$-coupling values calculated from MD simulations. Only Glu 7 could not be stereo-specifically assigned.

For the calculation of the backbone $^3J_{HN\text{-}H\alpha}$-couplings, the Karplus relation[4,5] was used with the parameter values a = 6.4 Hz, b = -1.4 Hz and c = 1.9 Hz[50], see Figure 2 of Ref.[26]. The side-chain $^3J_{H\alpha\text{-}H\beta}$-couplings were calculated using the parameter values a = 9.5 Hz, b = -1.6 Hz and c = 1.8 Hz[51], see Figure 2 of Ref.[26].

The experimentally derived $^3J_{HN\text{-}H\alpha}$-coupling values for Val 2, Thr 51, Asp 66, Cys 115, Thr 118 and Ile 124 lie outside the Karplus curve, so were set to 9.7 Hz, which is the maximum of the Karplus curve used[50]. None of the experimentally derived $^3J_{H\alpha\text{-}H\beta}$-coupling values lie outside the Karplus curve used[51]. The nomenclature for the $H_{\alpha 2}$ and $H_{\alpha 3}$ atoms in Gly residues and the H_β, $H_{\beta 2}$ and $H_{\beta 3}$ atoms in the side chains was defined as in Figure 3 of Ref.[52]. The values obtained from the simulations can be found in Tables S2 – S5 of Supporting Information.

In view of the various factors contributing to an uncertainty of about 2 Hz inherent to the Karplus relation linking structure and ^3J-couplings[1,6], a deviation of less than 2 Hz between ^3J-coupling values calculated from MD trajectory structures and ^3J-coupling values derived from experiment is considered insignificant.

3.5. S^2 order-parameter restraints for HEWL

Four sets of experimentally derived S^2 order-parameter for HEWL, 121 for the backbone NH and 79 for the side-chain CH_3, NH and NH_2 moieties[38,39], were used to evaluate the simulations, see Supporting Information Tables S6 – S9. S^2 order parameters for the atom pair (k_1,k_2) were calculated using the ensemble averaging expression Eq. (19).

For the Asn and Gln residues, only one S^2_{NH} *(exp)*-value per NH_2 group is available[38]. This required the assignment to one of the two NH1 and NH2 bond vectors. This was done based on a comparison of the S^2_{NH1} *(sim)*- and S^2_{NH2} *(sim)*-values calculated from MD simulations[27]. The experimentally unassigned S^2_{CG1}- and S^2_{CG2}-values for Val and S^2_{CD1}- and S^2_{CD2}-values for Leu residues[39] were assigned in a

similar way[27].

For an ideal methyl group with equal and fixed C-H bond lengths and H-C-H bond angles in which rotation around the symmetry axis occurs uniformly, the order parameter for the C-H bond vector is given by[53]

$$S^2_{rot} = ((3\cos^2(\beta)-1)/2)^2 \quad , \tag{21}$$

where β is the angle between a C-H vector and the symmetry axis, which can be considered equal to the C-C bond vector of the bond to the C-atom adjacent to the CH_3-group. When in addition the rotational motion around the C-C axis is independent of the motion of the C-axis itself, one may factorise their contributions,

$$S^2_{CH} = S^2_{CC}\, S^2_{rot} \ . \tag{22}$$

When β = 109.5°, one has[53] $S^2_{rot} = 0.111$. Thus the methyl group restraining is applied to the C-C bond vector and the target value is

$$S^2_{CC}\,(exp) = S^2_{CH}\,(exp)/0.111 \quad . \tag{23}$$

For the NH_2-groups in Asn and Gln approximation Eq. (22) does not hold, because the rotation around the C-N axis is not uniform. There is a large barrier for the 180° rotation and the rotational motion need not be decoupled from other motions. Experimentally, the two hydrogens are in slow exchange[38].

Order-parameter target values greater than 0.95 were set to 0.95. The S^2 order-parameter values obtained from the simulations can be found in Tables S6 – S9 of Supporting Information.

In view of the uncertainty due to the assumptions and approximations involved in the conversion of NMR relaxation data to $S^2_{k_1k_2}$ *(exp)*-values corresponding to

experiment and inherent to the calculation of $S^2_{k_1k_2}$ *(sim)*-values from MD simulations, a deviation of less than 0.2 between simulation and experiment is considered insignificant.

3.6. MD simulations performed

Four types of MD simulations of HEWL of length t^{MD} = 100 ns were performed resulting in trajectories or ensembles of HEWL structures.

1. MD simulation of HEWL without any restraining of the protein using the 54A7 force field (*MD_54A7*).
2. MD simulation of HEWL without any restraining of the protein using the 54A8 force field (*MD_54A8*).
3. MD simulation of HEWL with NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restraining using the 54A7 force field (*rMD_54A7*).
4. MD simulation of HEWL with NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restraining using the 54A8 force field (*rMD_54A8*).

3.7. Analysis of atomic trajectories

Trajectory energies and atomic coordinates were stored at 2 ps intervals and used for analysis[28,54].

Atom-positional root-mean-square differences RMSD between calculated structures and the *2VB1* X-ray crystal structure and atom-positional root-mean-square fluctuations (RMSF), i.e. around their average positions, in the MD trajectories were calculated after superimposing the backbone atoms (N, CA, C) of residues 3 – 126 to eliminate the contribution of overall translation and rotation of the protein.

The secondary structure assignment was done with the program DSSP, based on the Kabsch-Sander rules[55].

Hydrogen bonds were identified according to a geometric criterion: a hydrogen bond was assumed to exist if the hydrogen-acceptor distance was smaller than 0.25 nm and the donor-hydrogen-acceptor angle was larger than 135°.

4. Results and Discussion

Figure S1 shows the atom-positional root-mean-square difference (RMSD) for the backbone atoms of HEWL between the *2VB1* X-ray crystal structure and the MD simulated structures as function of time. After about 50 ns the RMSD value appears to be stable for the two unrestrained MD simulations, while for the restrained simulations stabilisation occurs earlier, after about 10 ns. As expected, applying 2043 restraints stabilises the distance from the X-ray structure of the protein. This suggests that averaging over 100 ns of the simulations would be sufficient to characterise the ensemble of configurations of HEWL in aqueous solution. The occurrence of secondary structure elements as function of time, shown in Figures S2 - S5, appears to confirm this conclusion.

4.1. Analysis of unrestrained MD simulations: comparison between force fields 54A7 and 54A8

A comparison of the occurrence of secondary structure elements as function of time (Figures S2 and S3) between the two force fields used shows overall similar behaviour. Small differences are observed at residue 78, where the 54A7 force field tends to indicate a 3_{10}-helical fold, and around residue 108, where the 54A7 force field indicates a 3_{10}-helical fold, whereas force field 54A8 seems to prefer an α-helical fold. At residue 116 the opposite is observed (54A7: α-helical; 54A8: 3_{10}-helical). The difference between these two force fields only concerns the partial charges of the charged side chains (Arg, Lys, Asp, Glu 7, His 15) and chain termini, leading to differences in electric fields throughout the protein, in particular close to these residues. This may

influence the balance between adopting a 3_{10}-helical and α-helical local structure in the backbone. Table 1 shows, in line with the previous observations, rather small differences in average potential energies in the *MD_54A7* and *MD_54A8* simulations.

Table 2 shows that force field 54A8 yields 53 NOE atom – atom distance upper bound violations larger than 0.1 nm (out of 1630 upper bounds, 3%), a bit more than the 34 upper bound violations found using the 54A7 force field (2%).

A similar picture is seen in Table 3 for the 95 backbone (assigned) $^3J_{HN\text{-}H\alpha}$ coupling constants, with 13 deviations larger than 2 Hz from measured values for force field 54A7 (14%) and 19 for force field 54A8 (20%). Of the 22 unassigned backbone $^3J_{HN\text{-}H\alpha}$ coupling constants only 2 (9%), for force field 54A8, deviate more than 2 Hz from the measured value, see Table 4. Table 5 shows that for the 58 (assigned) side-chain $^3J_{H\alpha\text{-}H\beta}$ coupling constants 14 (24%) deviate more than 2 Hz from the measured values, for the 54A7 as well as the 54A8 force field. For the 38 unassigned side-chain $^3J_{H\alpha\text{-}H\beta}$ coupling constants (Table 6) these values are 13 (34%) for force field 54A7 and 10 (26%) for force field 54A8.

Table 7 shows the number of deviations between 121 simulated and experimentally derived backbone S^2_{NH} order parameter values. For force field 54A7 17 values(14%) are larger than 0.2, while for force field 54A8 there are 15 values (12%) larger than 0.2. For the 51 side-chain S^2_{CH} order parameters the numbers of deviations larger than 0.2 are 18 for force field 54A7 (35%) and 21 (41%) for force field 54A8 (Table 8). For the 11 side-chain S^2_{NH} order parameters of Trp and Arg residues the numbers of deviations larger than 0.2 are 1 for force field 54A7 (9%) and 2 (18%) for force field 54A8 (Table

9). Of the 17 side-chain S^2_{NH2} order parameters of Asn and Gln residues there is only one deviation larger than 0.2 (6%) for force field 54A7 (Table 10).

Figure 2 shows the atom-positional root-mean-square fluctuations (RMSF) for the CA backbone atoms in the four simulations. The two force fields used, 54A6 and 54A7, show rather similar mobility in the unrestrained simulations.

Overall, the agreement between the 2043 MD simulated and experimentally derived data for HEWL is rather similar for both force fields.

4.2. Analysis of unrestrained MD simulations: comparison with the *2VB1* X-ray crystal structure

Although the *2VB1* X-ray crystal structure was not determined at the low pH of about 3.5 of the NMR measurements of HEWL in aqueous solution, it satisfies rather well the 1630 NOE atom – atom distance upper bounds, with only 12 violations (1%) larger than 0.1 nm and none larger than 0.2 nm (Table 2). The unrestrained MD simulations yield 34 (54A7, 2%) and 53 (54A8, 3%) bound violations larger than 0.1 nm and only 11 (54A7, 1%) and 18 (54A8, 1%) larger than 0.2 nm. The number of deviations from measured values for the 95 backbone (assigned)) $^3J_{HN\text{-}H\alpha}$ coupling constants larger than 2 Hz is 3 (3%) for the *2VB1* X-ray crystal structure, while 13 (54A7, 14%) and 19 (54A8, 20%) for the MD simulations (Table 3). For the 22 unassigned backbone $^3J_{HN\text{-}H\alpha}$ coupling constants (Table 4) the MD simulations show a similar agreement with measured values as the *2VB1* X-ray crystal structure (only 2 deviations larger than 2 Hz for the 54A8 force field).

For the 58 assigned side-chain (Table 5) and 38 unassigned (Table 6) side-chain

$^3J_{H\alpha\text{-}H\beta}$ coupling constants the comparison between the *2VB1* X-ray crystal structure and the MD simulations shows a different picture. The *2VB1* X-ray crystal structure shows 15 (26%) deviations from measured and assigned values larger than 2 Hz, while both MD simulations yield 14 (24%) such deviations, whereas for the unassigned $^3J_{H\alpha\text{-}H\beta}$ coupling constants these numbers are 25 (*2VB1* X-ray crystal structure, 66%), 13 (54A7, 34%) and 10 (54A8, 26%) for the MD simulations. It is not surprising that an X-ray crystal structure will match side-chain conformations less well than backbone ones.

The GROMOS force fields appear to match the NMR derived (1630) NOE atom – atom distance upper bounds and (95 assigned and 22 unassigned) backbone $^3J_{HN\text{-}H\alpha}$ coupling constants for HEWL less well than the *2VB1* X-ray crystal structure. This may have different origins.

1. The protein HEWL is a challenging case regarding detailed structure determination and refinement: it is a non-spherical, not compact protein with an overall charge of +10*e*, containing a variety of secondary structure elements and loops. As will be discussed below, the latter show considerable structural variability.

2. The GROMOS force-field parameter calibration protocol primarily involves reproduction of condensed-phase thermodynamic quantities such as energies, densities, and solvation free energies of small compounds representative for protein fragments. It is not based on (machine) learning from X-ray crystal structures for proteins.

The calibration protocol for the GROMOS force fields lead to simulated destabilisation properties of proteins matching experiment, thereby avoiding

overstabilisation of proteins, as other protein force fields seem to do. An example is the reproduction of the temperature and urea induced denaturation of the Trp-cage mini protein TC5b[56]. MD simulations of TC5b in aqueous solution at three different temperatures (278 K, 360 K, 400 K) show the melting temperature to be lower than 360 K, which is much closer to the experimental value of about 315 K than the values of 400 – 450 K reported for other force fields[56]. This flexibility of the GROMOS force fields would suggest that the deviations from the 2043 experimentally derived data observed in the unrestrained MD simulations would be easily reduced by restraining to these data.

4.3. Analysis of NOE, ^3J-coupling and S^2 order-parameter restrained MD simulations

Applying the 2043 NOE atom – atom distance, ^3J-coupling and S^2 order-parameter time-averaging restraints reduces the number of significant deviations from experimentally derived data to only a mere 3 (*rMD_54A7*) or 2 (*rMD_54A8*) for the 51 side-chain S^2_{CH} order-parameter values (Table 8). Table 1 shows the different restraining energies. We note that local-elevation ^3J-coupling restraining does not introduce strain into the structural ensemble in the same way as distance restraining does, because in local-elevation restraining the potential-energy surface is only adapted, on the fly, until the simulated ^3J-coupling values agree with the experimentally determined target ^3J-coupling values. The restrained MD simulations of HEWL in aqueous solution produce a structural ensemble that is compatible with the NMR data.

The regions that form 3_{10}-helices in the X-ray structures of HEWL (residues 80-84

and 120-124) have been of interest as these regions formed α-helices in the NMR-derived solution structure and some of the main chain $^{3}J_{HN\text{-}H\alpha}$ coupling constant values are higher than would be expected for a helix (7.2 and 9.2 Hz for residues 83 and 84 respectively)[24,37]. In the simulations the residues in these regions form both NH(i) – CO(i-3) and NH(i) – CO(i-4) hydrogen bonds (Table 11) with much lower populations (mostly 20-50%) than are seen for the hydrogen bonds in the α-helical regions (generally 80-95%). For residues 80-84 the 3_{10}-helical hydrogen bonds predominate in all simulations. For residues 120-123 the α-helical hydrogen bonds predominate in the unrestrained simulations, while α-helical and 3_{10}-helical hydrogen bonds have similar populations in the restrained simulations. Thus the simulations suggest that the 3_{10}-helices persist in solution but are fluctuating in nature and have some α-helical character too.

Various regions of the protein display more mobility and have lower experimentally derived ^{1}H-^{15}N order-parameter values and show significant root-mean-square fluctuations (RMSF) in the MD simulations (particularly residues 46-50, 68-72 and around residues 103 and 118), see Figures S4, S5 and 3. The hydrogen bonds of these residues have been analysed in detail in the MD simulations and compared with those in three X-ray structures of HEWL with different crystal symmetries (Table 11). The data for the region around Asn 103 are particularly interesting. For these residues there are differences in the hydrogen bonds seen in the different X-ray structures, some large NOE violations for the unrestrained 54A8 simulation and some $^{3}J_{HN\text{-}H\alpha}$ coupling-constant deviations for both unrestrained simulations. Here three of the hydrogen bonds

seen in the X-ray structures have very low populations in all the simulations. One is present only in the 54A7 unrestrained simulation and the other two are seen with fairly low populations in the two restrained simulations. This region therefore has fewer hydrogen bonds in solution than the X-ray structures would suggest.

Residues 68-72 in the long loop region also show some variation in the hydrogen bonds present in the different X-ray structures and show some $^{3}J_{HN\text{-}H\alpha}$ coupling constant deviations for both unrestrained simulations. Again some of the hydrogen bonds seen in the X-ray structure have very low populations in the simulations, but there are some hydrogen bonds with populations of 30-55% in all the simulations which are not seen in the X-ray structures (particularly 69 NH – 66 OD1/2 and 69 OHG1 – 66 OD1/2). Hence this region has a significant number of hydrogen bonds in solution but some of the predominant ones are not observed in the X-ray structures.

For residues 46-50 the main chain hydrogen bonds seen in the X-ray structures are observed with high populations (61-93%) in all the simulations, but there is more variation in the side chain hydrogen bonding. In a similar manner, in the region around residue 118 in the restrained simulations the main chain hydrogen bonds seen in the X-ray structures are observed, but there is also a hydrogen bond involving the side chain of Thr 118 (118 OHG1 – 115 CO) which is not seen in the X-ray structure.

Overall, in the regions of higher main-chain fluctuations a greater variety of hydrogen bonds is seen in the simulations than in the X-ray structures. Many of the X-ray hydrogen bonds involving side chains do not persist in solution.

5. Conclusions

When comparing the MD simulations of HEWL in aqueous solution using the slightly different GROMOS force fields 54A7 and 54A8 the structural and energetic differences are small. This is not surprising because these force fields only differ in regard to the partial charges of the charged side chains (Arg, Lys, Asp, Glu 7, His 15) and the chain termini, leading to differences in electric fields in throughout the protein. This may influence the balance between adopting 3_{10}-helical versus α-helical local structure in the backbone.

Both force fields show comparable agreement with the NMR data. Of the 1630 NOE atom – atom distance upper bounds 98% (54A7) and 97% (54A8) are reproduced within 0.1 nm. Of the 117 (95 stereo-specifically assigned and 22 unassigned) backbone $^3J_{HN\text{-}H\alpha}$ coupling constants 89% (54A7) and 82% (54A8) are reproduced within 2 Hz from the experimentally derived values. Of the 96 (58 stereo-specifically assigned and 38 unassigned) side-chain $^3J_{H\alpha\text{-}H\beta}$ coupling constants 72% (54A7) and 75% (54A8) are reproduced within 2 Hz from the experimentally derived values. Of the 121 experimentally derived S^2_{NH} backbone order-parameter values 86% (54A7) and 88% (54A8) are reproduced within 0.2. Of the 79 (51 S^2_{CH}, 11 S^2_{NH}, 17 S^2_{NH2}) side-chain S^2 order-parameter values 75% (54A7) and 71% (54A8) are reproduced within 0.2.

Applying the 2043 NOE atom – atom distance, ^3J-coupling and S^2 order-parameter restraints produce a structural ensemble for HEWL in aqueous solution that is compatible with the NMR data. Out of the 2043 restraints only a mere 3

(*rMD_54A7*) or 2 (*rMD_54A8*) of the simulated 51 side-chain S^2_{CH} order-parameter values deviate more than 0.2 from the experimentally derived values.

A comparison of the structural ensembles of HEWL in aqueous solution generated in the NMR-data restrained MD simulations with X-ray crystal structures show that the latter are generally compatible with the NMR data on the backbone, but less so with the ^3J-coupling NMR data on the side chains. These are better represented by the restrained MD simulations. Hydrogen bonding patterns in regions that form 3_{10}-helical or α-helical structure (residues 80 - 84 and 120 – 124), and around residues 46 – 50, residues 118 and 103 are different. In aqueous solution there seems to be a greater variety of hydrogen bonding networks than suggested by the X-ray crystal structures. Many of the hydrogen bonds involving side chains in the crystalline environment are not present in solution, which seems not unexpected, but suggests caution when interpreting side-chain conformations present in X-ray crystal structures.

This example of HEWL illustrates that a protein in aqueous solution may show considerably more structural variety than indicated by X-ray crystal structures derived from diffraction data.

Data and Software Availability

The GROMOS software for biomolecular simulation is available at https://github.com/biomos under the GPL-2.0 license.

Supporting Information

Table S1. List of 1630 NOE bounds and distances for HEWL calculated for the *2VB1* X-ray structure and the r^{-3} weighted averages of the MD simulations.

Tables S2 – S5. Two sets of backbone $^{3}J_{HN\text{-}H\alpha}$ couplings (95, 22) and two sets of side-chain $^{3}J_{H\alpha\text{-}H\beta}$ couplings (58, 38) for HEWL calculated for the *2VB1* X-ray structure and the averages of the MD simulations.

Tables S6 – S9. One set of backbone S^2_{NH} order parameters (121) and three sets of side-chain S^2_{CH} (51), S^2_{NH1} (11) and S^2_{NH2} (17) order parameters for HEWL calculated as averages over the MD simulations.

Figure S1. Atom-positional root-mean-square difference (RMSD) for the backbone atoms of HEWL between the *2VB1* X-ray crystal structure and the MD simulated structures as function of time.

Figures S2 – S5. Backbone secondary structure as function of time for the four MD simulations.

Acknowledgement

LJS acknowledges the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out some of this work,

http://dx.doi.org/10.5281/zenodo.22558. NH acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2075 – 390740016. Parts of the computations were performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and Arts and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

References

[1] W. F. van Gunsteren, J. A. Allison, X. Daura, J. Dolenc, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu, L. J. Smith, Deriving Structural Information from Experimentally Measured Data on Biomolecules, *Angew. Chem. Int. Ed.* **2016**, *55***,** 15990-16010.

[2] D. Neuhaus, M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, **2000**, 2nd edition, Wiley-VCH, New York.

[3] C. Peter, X. Daura, W. F. van Gunsteren, Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulation, *J. Biomol. NMR* **2001**, *20*, 297-310.

[4] M. Karplus, Contact electron–spin coupling of nuclear magnetic moments, *J. Chem. Phys.* **1959**, *30*, 11-15.

[5] M. Karplus, Vicinal proton coupling in nuclear magnetic resonance, *J. Amer. Chem. Soc.* **1963**, *85*, 2870-2871.

[6] J. R. Allison, W. F. van Gunsteren, A method to explore protein side chain conformational variability using experimental data, *ChemPhysChem* **2009,** *10*, 3213-3228.

[7] K. Park, W. F. van Gunsteren, Parameter Optimization for Calculation of Proton Chemical Shift in Protein, *J. Korean Mag. Res. Soc.* **1997**, *1*, 71-78.

[8] D. S. Wishart, Interpreting protein chemical shift data, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *58*, 62-87.

[9] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules, 1. Theory and Range of Validity, *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.

[10] E. R. Henry, A. Szabo, Influence of vibrational motion on solid state line shapes and

NMR relaxation, *J. Chem. Phys.* **1985**, *82*, 4753-4761.

[11] N. Hansen, F. Heller, N. Schmid, W. F. van Gunsteren, Time-averaged order parameter restraints in molecular dynamics simulations, *J. Biomol. NMR* **2014**, *60*, 169–187.

[12] M. Pechlaner, W. F. van Gunsteren, L. J. Smith, N. Hansen, Molecular Structure Refinement of a Beta-Heptapeptide Based on Residual Dipolar Couplings: The Challenge of Extracting Structural Information from Measured RDCs, *J. Phys. Chem. B*, **2025**, *129*, 3131-3158.

[13] M. Pechlaner, W. F. van Gunsteren, N. Hansen, L. J. Smith, On the Use of Measured Residual Dipolar Couplings to Calculate Residual Dipolar Couplings for a Protein Structure: A Case Study Using Hen Egg-White Lysozyme, *J. Chem. Inf. Model.,* **2025**, doi: 10.1021/acs.jcim.5c01428.

[14] D. Poger, W. F. van Gunsteren, A. E. Mark, A new force field for simulating phosphatidylcholine bilayers, *J. Comput. Chem.* **2010**, *31*, 1117-1125.

[15] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, W. F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, *Eur. Biophys. J.* **2011**, *40*, 843-856.

[16] M. M. Reif, P. H. Hünenberger, C. Oostenbrink, New Interaction Parameters for Charged Side Chains in the GROMOS Force Field, *J. Chem. Theory Comput.* **2012**, *8*, 3705-3723.

[17] P. Jolles, J. Jolles, What's new in lysozyme research—always a model system, today as yesterday? *Mol. Cell Biochem*. **1984**, *63*, 165–189.

[18] C. M. Dobson, P. A. Evans, S. E. Radford, Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **1994**, *19*, 31-37.

[19] R. Swaminathan, V. K. Ravi, S. Kumar, M. V. S. Kumar, N. Chandra, Lysozyme: A model protein for amyloid research. *Advances in Protein Chemistry and Structural Biology* **2011**, *84*, 63-111.

[20] D. J. Vocadlo, G. J. Davies., R. Laine, S. G. Withers, Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **2001**, *412*, 835–838.

[21] H. Wei, Z. Wang, J. Zhang, S. House, Y. -G. Gao, L. Yang, H. Robinson, L. H. Tan, H. Xing, C. Hou, I. M. Robertson, J. -M. Zou,, L. Yu, Time-dependent, protein-directed growth of gold nanoparticles within a single crystal of lysozyme. *Nature Nanotech* **2011**, *6*, 93–97.

[22] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. T. C. North, D. C. Philips, V. R. Sarma, Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **1965**, *206*, 757–761.

[23] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, Structure of Hen Lysozyme in Solution, *J. Mol. Biol.* **1993**, *229*, 930–944.

[24] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, A refined solution structure of hen lysozyme determined using residual dipolar coupling data *Protein Science*, **2001**, *10*, 677-688.

[25] L. J. Smith, W. F. van Gunsteren, N. Hansen, On the Use of Time-Averaging Restraints when Deriving Biomolecular Structure from ^{3}J-coupling Values Obtained from NMR Experiments, *J. Biomol. NMR* **2016**, *66*, 69-83.

[26] L. J. Smith, W. F. van Gunsteren, N. Hansen, On the Use of ^{3}J-coupling NMR Data to Derive Structural Information on Proteins, *J. Biomol. NMR* **2021**, *75*, 39-70.

[27] L. J. Smith, W. F. van Gunsteren, N. Hansen, On the Use of Side-Chain NMR Relaxation

Data to Derive Structural and Dynamical Information on Proteins: A Case Study Using Hen Lysozyme, *ChemBioChem* **2021**, ***22***, 1049-1064.

[28] W. F. van Gunsteren et al., *The GROMOS Software for (Bio)Molecular Simulation. Volumes 1 – 9,* **2024**, http://www.gromos.net/. Accessed 4 January 2024.

[29] W. F. van Gunsteren, R. Boelens, R. Kaptein, R. M. Scheek, E. R. P. Zuiderweg, An Improved Restrained Molecular Dynamics Technique to Obtain Protein Tertiary Structure from Nuclear Magnetic Resonance Data, In: "Molecular Dynamics and Protein Structure", J. Hermans ed., Polycrystal Book Service, P.O. Box 27, Western Springs, Ill. 60558, USA, **1985**, 92-99.

[30] 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, Switzerland, **1996**, 1-1042, ISBN 3 7281 2422 2.

[31] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, Time-dependent distance restraints in molecular dynamics simulations, *Chem. Phys. Lett.* **1989**, *157*, 289 - 294.

[32] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, Time-averaged Nuclear Overhauser Effect Distance Restraints Applied to Tendamistat, *J. Mol. Biol.* **1990**, *214*, 223-235.

[33] W. R. P. Scott, A. E. Mark, W. F. van Gunsteren, On using time-averaging restraints in molecular dynamics simulation, *J. Biomol. NMR* **1998**, *12*, 501-508.

[34] K. Wüthrich, M. Billeter, W. Braun, 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.* **1983**, *169*, 949–961.

[35] A. P. Nanzer, A. E. Torda, C. Bisang, C. Weber, J. A. Robinson, W. F. van Gunsteren, Dynamical Studies of Peptide Motifs in the *Plasmodium falciparum* Circumsporozoite Surface Protein by Restrained and Unrestrained MD Simulations, *J. Mol. Biol.* **1997**, *267*, 1012-1025.

[36] T. Huber, A. E. Torda, W. F. van Gunsteren, Local elevation: A method for improving the searching properties of molecular dynamics simulation, *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695-708.

[37] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, Analysis of φ and χ_1 torsion angles for hen lysozyme in solution from ^{1}H NMR spin-spin coupling constants, *Biochemistry* **1991**, *30*, 986-996.

[38] M. Buck, J. Boyd, C. Redfield, D. A. MacKenzie, D. J. Jeenes, D. B. Archer, C. M. Dobson, Structural Determinants of Protein Dynamics : Analysis of ^{15}N NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg White Lysozyme, *Biochemistry* **1995**, *34*, 4041-4055.

[39] V. Moorman, K. G. Valentine, A. J. Wand, The dynamical response of hen egg white lysozyme to the binding of a carbohydrate ligand, *Prot. Sci.* **2012**, *21*, 1066-1073.

[40] K. Bartik, C. Redfield, C. M. Dobson, Measurement of the individual pK_a values of acidic residues of Hen and Turkey Lysozymes by two-dimensional ^{1}H NMR. *Biophys. J.* **1994**, *66*, 1180-1184.

[41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, **1981**, in *Intermolecular forces*, B. Pullmann Ed., Reidel, Dordrecht, pp 331-342.

[42] J. –P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys.* **1977**, *23*, 327-341.

[43] R. W. Hockney, J. W. Eastwood, *Computer Simulation Using Particles*, **1981**, McGraw-Hill, New York.

[44] W. F. van Gunsteren, H. J. C. Berendsen, R. G. Geurtsen, H. R. J. Zwinderman, A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: Comparison with experimental two-dimensional NMR data, *Ann. New York Acad. Sci.* **1986**, *482*, 287-303.

[45] J. A. Barker, R. O. Watts, Monte Carlo studies of the dielectric properties of water-like models, *Mol. Phys.* **1973**, *26*, 789-792.

[46] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, A generalized reaction field method for molecular dynamics simulations, *J. Chem. Phys.* **1995**, *102*, 5451-5459.

[47] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations, *J. Chem. Phys.* **2001**, *115*, 1125-1136.

[48] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, www.pdb.org. *Nucleic Acids Res.* **2000**, *28*, 235-242.

[49] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **1984**, *81*, 3684-3690.

[50] A. Pardi, M. Billeter, K. Wüthrich, Calibration of the angular dependence of the amide

proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure, *J. Mol. Biol.* **1984**, *180*, 741–751.

[51] A. deMarco, M. Llinás, K. Wüthrich, Analysis of the ^{1}H-NMR spectra of ferrichrome peptides. I. The non-amide protons, *Biopolymers* **1978**, *17*, 617-636.

[52] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, Recommendations for the presentation of NMR structures of proteins and nucleic acids, *J. Biomol. NMR* **1998**, *12*, 1–23.

[53] D. C. Chatfield, A. Szabo, B. R. Brooks, Molecular Dynamics of Staphylococcal Nuclease: Comparison of Simulation with 15N and 13C NMR Relaxation Data, *J. Am. Chem. Soc.* **1998**, *120*, 5301-5311.

[54] A. P. Eichenberger, J. R. Allison, J. Dolenc, D. P. Geerke, B. A. C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang, W. F. van Gunsteren, GROMOS++ software for the analysis of biomolecular simulation trajectories, *J. Chem. Theory. Comp.* **2011**, *7*, 3379-3390.

[55] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* **1983**, *22*, 2577–2637.

[56] Z. Gattin, S. Riniker, P. J. Hore, K. H. Mok, W. F. van Gunsteren, Temperature and urea induced denaturation of the TRP-cage mini protein TC5b: a simulation study consistent with experimental observations, *Protein Science* **2009**, *18*, 2090-2099.

Tables

Table 1. Average energies (kJmol^{-1}) of the protein HEWL in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^{3}J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure. [a]Not including the restraining energies.

Type of energy (kJmol^{-1})	*MD_54A7*	*MD_54A8*	*rMD_54A7*	*rMD_54A8*
Potential energy[a]	-9893	-9754	-8730	-8695
Fluctuation of potential energy	275	272	292	338
Bond-angle energy	2347	2373	2410	2424
Improper-dihedral energy	700	696	743	745
Proper-dihedral energy	1340	1323	1532	1524
Electrostatic energy	-10875	-10779	-10186	-10210
Van der Waals energy	-3405	-3368	-3229	-3178
Restraining energy			477	514
NOE distance-restraining energy			107	109
^{3}J-coupling-restraining energy			335	370
S^2 order-parameter-restraining energy			35	35

Table 2. Number of NOE distance bound violations in the *2VB1* X-ray crystal structure (*X-ray_2VB1*), in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_4A8*) starting from the *2VB1* X-ray crystal structure. Number of NOE distance bounds: 1630.

Structure or simulation	Size of NOE distance bound violation (in nm)					
	0.05 – 0.1	0.1 – 0.15	0.15 – 0.2	0.2 – 0.25	0.25 – 0.3	> 0.3
X-ray_2VB1	21	7	5	0	0	0
MD_54A7	43	14	9	1	5	5
MD_54A8	45	19	16	9	5	4
rMD_54A7	1	0	0	0	0	0
rMD_54A8	1	0	0	0	0	0

Table 3. Number of deviations, $|^3J_{HN\text{-}H\alpha}$ (exp) - $^3J_{HN\text{-}H\alpha}$ (MD or X-ray)|, for the 95 backbone $^3J_{HN\text{-}H\alpha}$-coupling values derived and assigned based on NMR measurements (set *bb1*), in the *2VB1* X-ray crystal structure, in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Crystal structure or simulation	Size of $^3J_{HN\text{-}H\alpha}$ deviation (in Hz)				
	1 – 2	2 – 3	3 – 4	4 – 5	> 5
X-ray_2VB1	13	2	1	0	0
MD_54A7	32	7	6	0	0
MD_54A8	25	12	7	0	0
rMD_54A7	1	0	0	0	0
rMD_54A8	1	0	0	0	0

Table 4. Number of deviations, $|^3J_{HN\text{-}H\alpha}$ (exp) - $^3J_{HN\text{-}H\alpha}$ (MD or X-ray)|, for the 22 backbone $^3J_{HN\text{-}H\alpha}$-coupling values derived but stereo-specifically unassigned from NMR measurements (set *bb2*), in the *2VB1* X-ray crystal structure, in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Crystal structure or simulation	Size of $^3J_{HN\text{-}H\alpha}$ deviation (in Hz)				
	1 – 2	2 – 3	3 – 4	4 – 5	> 5
X-ray_2VB1	5	1	0	0	0
MD_54A7	3	0	0	0	0
MD_54A8	2	2	0	0	0
rMD_54A7	0	0	0	0	0
rMD_54A8	0	0	0	0	0

Table 5. Number of deviations, $|^3J_{H\alpha\text{-}H\beta}$ (exp) - $^3J_{H\alpha\text{-}H\beta}$ (MD or X-ray)|, for the 58 side-chain $^3J_{H\alpha\text{-}H\beta}$-coupling values derived and stereo-specifically assigned based on NMR measurements (set *sc1*), in the *2VB1* X-ray crystal structure, in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Crystal structure or simulation	Size of $^3J_{H\alpha\text{-}H\beta}$ deviation (in Hz)				
	1 – 2	2 – 3	3 – 4	4 – 5	> 5
X-ray_2VB1	23	9	4	1	1
MD_54A7	19	3	5	4	2
MD_54A8	20	6	2	2	4
rMD_54A7	3	0	0	0	0
rMD_54A8	6	0	0	0	0

Table 6. Number of deviations, $|^3J_{H\alpha\text{-}H\beta}(\text{exp}) - {}^3J_{H\alpha\text{-}H\beta}(\text{MD or X-ray})|$, for the 38 side-chain $^3J_{H\alpha\text{-}H\beta}$-coupling values derived but stereo-specifically unassigned from NMR measurements (set *sc2*), in the *2VB1* X-ray crystal structure, in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Crystal structure or simulation	Size of $^3J_{H\alpha\text{-}H\beta}$ deviation (in Hz)				
	1 – 2	2 – 3	3 – 4	4 – 5	> 5
X-ray_2VB1	9	3	4	5	13
MD_54A7	6	8	3	1	1
MD_54A8	12	8	1	0	1
rMD_54A7	2	0	0	0	0
rMD_54A8	1	0	0	0	0

Table 7. Number of deviations, $|S^2(\text{exp}) - S^2(\text{MD})|$, for the 121 backbone S^2_{NH}-values in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Simulation	Size of S^2 deviation					
	0.05 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5	> 0.5
MD_54A7	21	44	15	2	0	0
MD_54A8	31	36	12	3	1	0
rMD_54A7	62	30	0	0	0	0
rMD_54A8	60	34	0	0	0	0

Table 8. Number of deviations, $|S^2(\text{exp}) - S^2(\text{MD})|$, for the 51 side-chain S^2_{CH}-values in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Simulation	Size of S^2 deviation					
	0.05 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5	> 0.5
MD_54A7	7	16	9	7	1	1
MD_54A8	9	12	9	8	2	2
rMD_54A7	13	18	3	0	0	0
rMD_54A8	15	16	2	0	0	0

Table 9. Number of deviations, $|S^2(\text{exp}) - S^2(\text{MD})|$, for the 11 side-chain S^2_{NH}-values of Trp and Arg residues in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Simulation	Size of S^2 deviation					
	0.05 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5	> 0.5
MD_54A7	4	4	1	0	0	0
MD_54A8	3	4	2	0	0	0
rMD_54A7	4	2	0	0	0	0
rMD_54A8	4	1	0	0	0	0

Table 10. Number of deviations, $|S^2(\text{exp}) - S^2(\text{MD})|$, for the 17 side-chain S^2_{NH2}-values of Asn and Gln residues in the two (force fields 54A7 and 54A8) unrestrained MD simulations (*MD_54A7, MD_54A8*), and in the two corresponding NOE atom-atom distance, ^3J-coupling and S^2 order-parameter restrained 100 ns MD simulations (*rMD_54A7, rMD_54A8*) starting from the *2VB1* X-ray crystal structure.

Simulation	Size of S^2 deviation					
	0.05 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5	> 0.5
MD_54A7	3	6	0	1	0	0
MD_54A8	4	8	0	0	0	0
rMD_54A7	5	3	0	0	0	0
rMD_54A8	6	5	0	0	0	0

Table 11. Hydrogen bonds as observed in three (*2VB1*: triclinic, *1IEE*: tetragonal, *1AKI*: orthorhombic) X-ray crystal structures and in 100 ns unrestrained and NMR-data restrained MD simulations of HEWL in aqueous solution using the GROMOS force fields 54A7 and 54A8. Hydrogen bonds were identified according to a geometric criterion: a hydrogen bond was assumed to exist if the hydrogen-acceptor distance was smaller than 0.25 nm and the donor-hydrogen-acceptor angle was larger than 135°.

Hydrogen bond	X-ray	*MD _54A7*	*MD _54A8*	*rMD _54A7*	*rMD _54A8*
82 NH – 79 CO	(100,-,-)	49	42	46	54
83 NH – 79 CO	(-,-,-)	10	14	9	9
83 NH – 80 CO	(100,-,-)	58	48	55	61
84 NH – 80 CO	(-,-,-)	11	22	16	13
84 NH – 81 CO	(100,-,-)	44	37	53	53
85 NH – 81 CO	(-,-,-)	12	11	19	17
85 NH – 82 CO	(100,-,-)	35	29	21	21
122 NH – 119 CO	(100,-,-)	6	6	5	7
123 NH – 119 CO	(-,-,-)	39	35	37	42
123 NH – 120 CO	(100,-,-)	11	14	20	19
124 NH – 120 CO	(-,-,-)	80	84	34	31
124 NH – 121 CO	(100,-,-)	2	2	27	31
125 NH – 121 CO	(100,-,-)	64	33	30	32
125 NH – 122 CO	(100,-,-)	9	8	29	27
102 NH – 98 CO	(-,-,-)	27	-	26	17
102 NH – 99 CO	(100,-,-)	3	-	1	1
103 NH – 101 OD1	(-,-,-)	-	14	-	-
103 NDH2 – 98 CO	(100,-,-)	-	-	-	-
104 NH- 101 OD1	(-,-,-)	52	-	-	-
104 NH- 102 CO	(-,-,-)	-	-	13	14
105 NH – 23 OH	(100,100,100)	1	2	1	10
105 NH – 101 CO	(-,-,-)	1	90	-	-

105 NH – 103 CO	(-,-,-)	47	-	-	1
106 NH – 102 CO	(-,-,-)	1	67	-	-
106 NH – 103 CO	(-,100,100)	3	5	27	27
106 NHD2 – 103 CO	(-,100,100)	2	-	12	12
67 NH – 65 OD1	(100,100,100)	1	3	20	30
68 NH – 66 OD1	(100,100,100)	26	24	36	46
68 NH – 66 OD2	(-,-,-)	25	28	43	40
68 NH22 – 51 OG1	(-,100,-)				
69 NH – 66 OD1	(-,-,-)	40	44	52	49
69 NH – 66 OD2	(-,-,-)	45	52	55	43
69 OHG1 – 60 OG	(100,100,100)	1	1	-	-
69 OHG1 – 66 OD1	(-,-,-)	26	42	52	62
69 OHG1 – 66 OD2	(-,-,-)	32	52	55	60
72 NH – 69 CO	(-,100,-)	3	2	2	1
73 NH – 61 CO	(100,100,100)	74	86	88	90
45 NH12 – 49 CO	(-100,-)	5	9	1	1
46 NH – 50 CO	(100,100,100)	83	89	91	92
46 ND2H – 50 OG	(100,100,-)	8	15	62	57
46 ND2H – 52 OG1	(100,-,-)	55	54	49	52
46 ND2H – 52 OG2	(-,100,-)	55	51	42	45
49 NH – 46 CO	(100,100,100)	76	81	72	78
50 NH – 48 OD1	(100,100,100)	43	50	30	41
50 NH – 48 OD2	(100,100,100)	40	48	34	40
116 NH – 111 CO	(100,100,100)	7	2	47	47
118 NH – 115 CO	(100,100,100)	1	1	29	23
118 NH – 116 CO	(-,-,-)	19	22	7	5
118 OHG1 – 115 CO	(-,-,-)	3	-	36	34

Figures

Figure 1. Ribbon picture of the structure of HEWL showing in colour the regions of particular interest. 3_{10} helices: residue 80-84 (red) and 120-124 (green) and more mobile regions: residues 46-50 (blue), 68-72 (magenta), 102-105 (cyan) and 116-118 (orange).

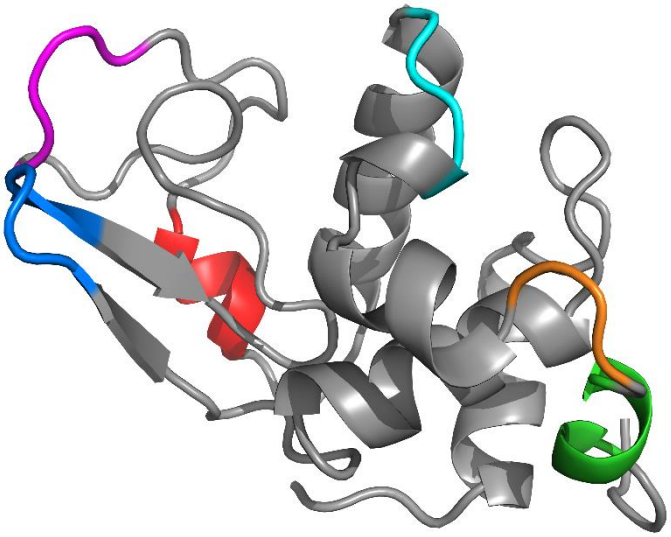


Figure 2. Atom-positional root-mean-square fluctuations (RMSF) for CA backbone atoms for the four 100 ns MD simulations, calculated after superimposing the backbone atoms (N, CA, C) of residues 3-126.

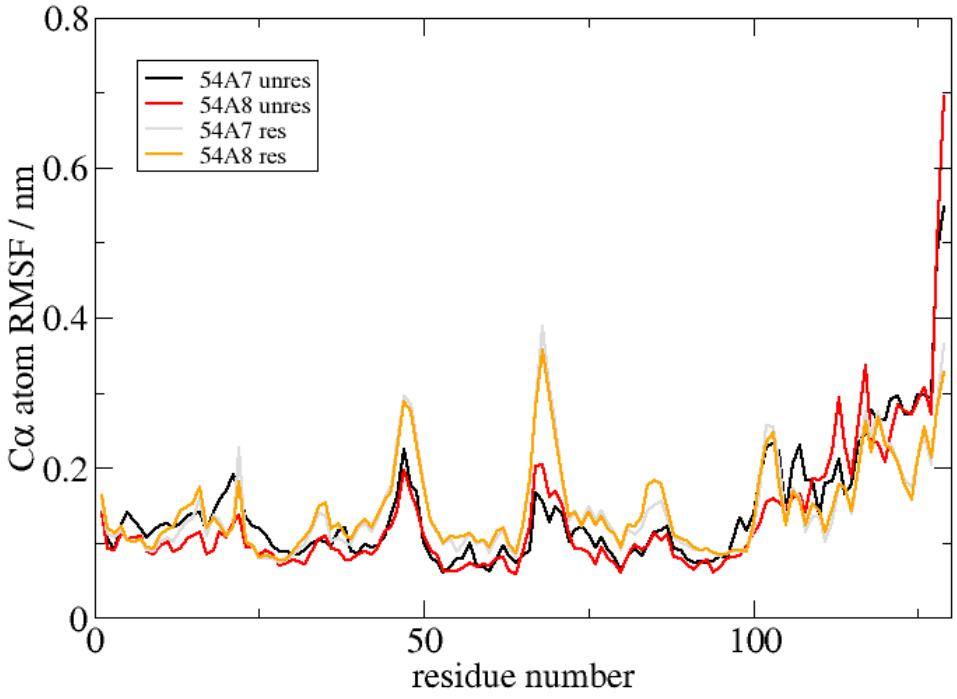
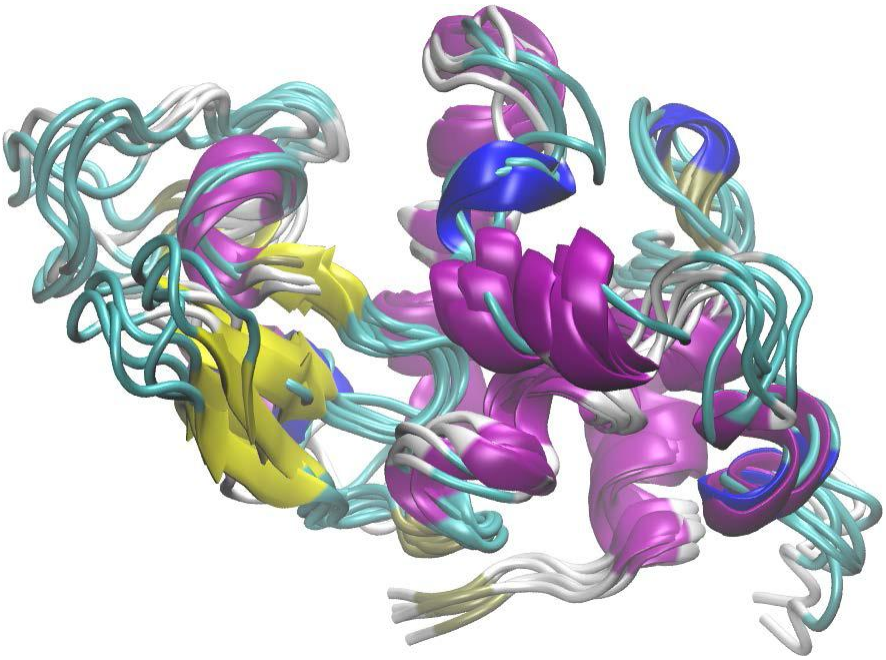


Figure 3. Six structures from the *rMD_54A7* trajectory (at 50, 60, 70, 80, 90, 100 ns) superimposed using the backbone atoms (N, CA, C) of residues 3 – 126 illustrating the variety of the local backbone and side-chain structures in aqueous solution.



TOC graphic