On the Use of Side-Chain NMR Relaxation Data to Derive Structural and Dynamical Information on Proteins: A Case Study Using Hen Lysozyme

Lorna J. Smith,[a] Wilfred F. van Gunsteren,[b] and Niels Hansen*[c]

Values of $S_{CH}$ and $S_{NH}$ order parameters derived from NMR relaxation measurements on proteins cannot be used straightforwardly to determine protein structure because they cannot be related to a single protein structure, but are defined in terms of an average over a conformational ensemble. Molecular dynamics simulation can generate a conformational ensemble and thus can be used to restrain $S_{CH}$ and $S_{NH}$ order parameters towards experimentally derived target values $S_{CH}^{(exp)}$ and $S_{NH}^{(exp)}$. Application of $S_{CH}$ and $S_{NH}$ order-parameter restraining MD simulation to bond vectors in 63 side chains of the protein hen egg white lysozyme using 51 $S_{CH}^{(exp)}$ target values and 28 $S_{NH}^{(exp)}$ target values shows that a conformational ensemble compatible with the experimentally derived data can be obtained by using this technique. It is observed that $S_{CH}$ order-parameter restraining of C–H bonds in methyl groups is less reliable than $S_{NH}$ order-parameter restraining because of the possibly less valid assumptions and approximations used to derive experimental $S_{CH}^{(exp)}$ values from NMR relaxation measurements and the necessity to adopt the assumption of uniform rotational motion of methyl C–H bonds around their symmetry axis and of the independence of these motions from each other. The restrained simulations demonstrate that side chains on the protein surface are highly dynamic. Any hydrogen bonds they form and that appear in any of four different crystal structures, are fluctuating with short lifetimes in solution.

1. Introduction

During the past 50 years, the determination of structure of proteins in crystal based on the reflections of X-rays has become a standard procedure to obtain information on proteins at the atomic level of resolution. Over the past 30 years NMR measurement of proteins in solution has also become a standard technique, not only to obtain information on protein structure, but also on dynamics at the atomic level of resolution. Several quantities that are observable by NMR, such as nuclear Overhauser enhancements (NOEs), $J$ couplings or chemical shifts, only provide local structural information. Only residual dipolar couplings (RDCs) do provide longer-range information in terms of the average (relative) directions of bond vectors throughout a molecule. In contrast, X-ray diffraction of crystals yields non-local information and, in addition, its 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, is much higher than that of NMR experiments on proteins in solution. On the other hand, NMR measurements may provide dynamic information in the form of relaxation data, for example, expressed as $S^2$ order parameters, and of all techniques available to obtain information on proteins in solution, NMR shows the highest information density.

All techniques to derive structural information from the measurement of observable quantities $Q$ make use of a relation of $Q$ to structure $r$, a function $Q(r)$. Since virtually all experimental techniques measure an average $\langle Q \rangle_{\text{space/time}}$ of $Q$ over the molecules (space) in the test tube and over a time window determined by the type of experiment, the derivation of structural information from a set of $Q$ values should account for the averaging involved in the measurement. This can be done by applying multi-molecule averaging or by time-averaging structure refinement instead of the commonly used single-structure refinement technique. Application of time-averaging structure refinement to proteins based on X-ray data,[4,5] NMR NOE[6,7] or $J$ coupling[8] data showed the protein structural variation to be much larger than that observed using single-structure refinement techniques.[9]

For observable quantities $Q$, such as X-ray reflection intensities $I_{\text{hkl}}$, NOEs (when represented as atom-atom distance bounds), $J$ couplings or chemical shifts, it is possible to formulate a function $Q(r)$ relating a $Q$ value to a particular...
structure. RDCs may be directly related to structure by assuming a single alignment tensor representing the anisotropic rotational distribution of the molecule, which is, unfortunately, unknown. For other observable quantities, such as $S^2$ order parameters, the function relating $Q$ to $r$ involves some average over the Boltzmann ensemble of structures in solution, $Q(\langle r(\tau) \rangle)$, where $f$ denotes the function of $r$ that is being averaged. This means that structure refinement based on such quantities must involve the averaging $\langle Q(\langle r(\tau) \rangle) \rangle$ in addition to the averaging $\langle Q(\langle r(f) \rangle) \rangle$.

$S^2$ order parameters are commonly derived from an analysis of NMR relaxation data using a so-called “model free” model[10] and can be calculated as an ensemble- or time-average of a function of the three Cartesian coordinate components of the $^{13}$C–$^1$H or $^{15}$N–$^1$H bond vector.[11] They are commonly interpreted as a measure of the geometrical restriction ($S^2 = 0$: no restriction; $S^2 = 1$: complete restriction) of the directions of the mentioned bond vectors on a time scale faster than the stochastic rotational tumbling of the molecule in solution, for proteins of the order of nanoseconds. For relatively stable proteins, order parameters involving backbone atoms lie generally in the range 0.75 < $S^2$ < 0.95. Exceptions occur in longer loops connecting secondary structure elements ($S^2$ values as low as 0.5) and for thermodynamic conditions, such as low pH, that are known to often destabilise proteins. For bond vectors in side chains, values as low as 0.05 can be found.

MD simulations have reproduced experimentally derived order-parameter values for bond vectors involving backbone atoms with some success.[12,13] For side chains, this is more challenging.[14–17] because of the flexibility of side chains and of the multiple hydrogens bound to the $^{13}$C atom in a CH$_3$ -group or bound to the $^{15}$N atom in a NH$_2$ -group. This suggests the use of $S^2$ order-parameter structure refinement using time averaging in order to obtain a conformational ensemble compatible with the order-parameter data. The technique of time-averaging order-parameter structure refinement has been tested on the backbone $^{15}$N–$^1$H order parameters of the B3 domain of protein G,[18] and subsequently applied to backbone $^{15}$N–$^1$H order parameters of the protein IL-4 at pH 6 to detect inconsistencies and model flaws regarding complementary sets of NMR data.[19] and applied to backbone $^{15}$N–$^1$H order parameters of the protein hGH at pH 2.7 in order to explain the occurrence of low order-parameter values in the middle of stable helices.[19]

The application of $S^2$ order-parameter restraining to CH$_3$ and NH$_2$ moieties in protein side chains is more challenging than to backbone NH groups. The multiple hydrogens cause ambiguity regarding peak assignments, which requires additional assumptions. The directional variability of bond vectors in side chains is generally larger than for backbone N–H or C–H vectors, leading to a greater variety of smaller order-parameter values. Third, averaging over side-chain motions may take longer to converge than over limited backbone motion in a stable protein.

Here the application of $S^2$ order-parameter restraining to side-chain NH, NH$_2$ and CH$_3$ moieties of the protein hen egg white lysozyme (HEWL) in investigation. An earlier comparison of the experimentally derived $S^2$ values with those obtained from short, 1 ns unrestrained MD simulations showed a poor relation between simulation and experiment,[16] which could be due to the short simulation time period or force-field deficiencies, assuming no flaws in the experimental data. Use of an improved force field, of much longer sampling, and of $S^2$ order-parameter restraining might generate a conformational ensemble compatible with these and other experimental data on HEWL.

HEWL is one of the proteins most studied. Several X-ray crystal structures are available, and sets of NOE data[22] of $J^\alpha$ couplings,[23] RDC values,[24] and of $S^2$ order parameters,[20,21] measured at a variety of thermodynamic conditions. Here the configurational ensembles, generated in unrestrained and in order-parameter restrained MD simulations, will be used to calculate side-chain $^{13}$C–$^1$H and $^{15}$N–$^1$H order-parameter values, which are compared to values obtained from NMR measurements at pH 3.5.[20,21]

The side-chain $S^2$ values were separated into two groups, one of C–H values and another of N–H values. By separately restraining these subsets of $S^2$ values, it could be investigated whether restraining one subset would improve the agreement with values derived from experiment for the other subset of $S^2$ values. Four X-ray crystal structures were used in the simulations and for comparison, in order to delineate the influence of a particular starting structure on the generated configurational ensemble.

The simulated configurational ensembles were also used to calculate NOE atom-atom distances that were compared to a set of NOE atom-atom distance bounds derived from experiment at pH 3.8.[22] A set of side-chain $J^\beta_{\text{NH}}$ couplings derived from experiment is available.[22,23] The 26 $J^\beta_{\text{NH}}$ couplings in side chains for which $S^2$ order-parameter values derived from experiment are available, were used for comparison. They regard $\chi_1$ torsional angles close to the backbone. A comparison of simulated with experimentally derived $J^\beta_{\text{NH}}$ coupling values is not straightforward though, because the Karplus relation $J^\beta(\theta)$ that connects a torsional angle $\theta$ to a $J^\beta$ coupling, possesses a rather large uncertainty of 1–2 Hz[23] small $J^\beta$ coupling values ($\approx 4$ Hz) are difficult to determine precisely from spectra, and $J^\beta$ couplings in the range 5–8 Hz may result from averaging over long time periods (microseconds). The set of RDCs for side chains of HEWL[24] was not used for comparison, because they strongly depend on the solvent environment in the measurement.[26]

Computational Methods

Energy minimisations and molecular dynamics simulations were performed using the GROMOS bio-molecular simulation software.[27–28]

Molecular model

The protein was modelled using the GROMOS bio-molecular force field 54 A7[16,21] In view of the pH used in the experimental NMR measurements, pH 3.5, only Glu35 was protonated and His was...
doubly protonated. The simple point charge (SPC) model was used to describe the solvent molecules in the rectangular periodic box. To compensate for the overall positive charge of the protein, 10 Cl⁻ ions were included in the solution. 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, allowing for a 2 fs MD time step in the leap-frog algorithm used to integrate the equations of motion. For the non-bonded interactions a triple-range method with cut-off radii of 0.8/1.4 nm was used. Short-range van der Waals and electrostatic interactions were evaluated every time step based on a charge-group pair list. 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 a reaction-field approximation with a relative dielectric permittivity of 61 was used. Minimum-image periodic boundary conditions were applied.

**Simulation set-up**

Four X-ray crystal structures were used as initial structures for the energy minimisations followed by MD simulations.

1. Structure 2VB1 from the Protein Data Bank (PDB), derived from a triclinic unit cell at 0.065 nm resolution at T = 100 K. It contains multiple side-chain conformations for 46 residues.
2. Structure 4LZT from the PDB, derived from a triclinic unit cell at 0.095 nm resolution at T = 295 K. It contains multiple side-chain conformations for 46 residues.
3. Structure 1IEE from the PDB, derived from a tetragonal unit cell at 0.110 nm resolution at T = 298 K. It contains multiple side-chain conformations for 33 residues.
4. Structure 1AKI from the PDB, derived from an orthorhombic unit cell at 0.15 nm resolution at T = 298 K. It contains no multiple side-chain conformations.

For the initial structures the side-chain conformation with the highest occupancy was chosen. An 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 boxes with 12157 water molecules for the initial protein structures. 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 with a force constant of 25000 kJ mol⁻¹ nm⁻².

The resulting protein-water configuration (i.e., coordinates) was used as initial configuration for the MD simulation. In order to avoid artificial deformations in the protein structure due to relatively high-energy atomic interactions still present in the system, the MD simulation was 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, 25000, 25000, and 25 kJ mol⁻¹ nm⁻². The temperature was kept constant using the weak coupling algorithm with a coupling time τ<sub>c</sub> = 0.5 ps and an isothermal compressibility K<sub>T</sub> = 4.575 × 10⁻⁴ (kJ mol⁻¹ nm⁻²). The centre of mass motion of the system was removed every 1000 time steps (2 ps).

**Order-parameter restraining**

Two sets of ¹³C—H and ¹⁵N—H side-chain order-parameter target values S<sup>(exp)</sup><sub>CH</sub> and S<sup>(exp)</sup><sub>NH</sub> were used, see Tables 1–3.

1. A set of 51 S<sup>(exp)</sup><sub>CH</sub> values for CH<sub>1</sub> moieties in 30 residues.
2. A set of 28 S<sup>(exp)</sup><sub>NH1</sub> and S<sup>(exp)</sup><sub>NH2</sub> values for NH moieties in six Trp, five Arg, fourteen Asn and three Gln residues.

The distribution of these S<sup>(exp)</sup> values over the protein is shown in Figure 1. For the Asn and Gln residues, the S<sup>(exp)</sup><sub>CH</sub> value per NH<sub>2</sub> group was available. This required the assignment to one of the two NH1 and NH2 bond vectors. This was done by calculating S<sup>2</sup><sub>CH</sub>(MD) and S<sup>2</sup><sub>CH</sub>(MD) values from the unrestrained simulation MD<sub>2VB1</sub> starting from the 2VB1 X-ray structure and then selecting the N–H vector with its S<sup>2</sup><sub>CH</sub>(MD) value closest to S<sup>(exp)</sup><sub>CH</sub> for restraining. A corresponding procedure was used to assign experimentally unassigned S<sup>(exp)</sup><sub>CG1</sub> and S<sup>(exp)</sup><sub>CG2</sub> values for Val residues and S<sup>(exp)</sup><sub>CG1</sub> and S<sup>(exp)</sup><sub>CG2</sub> values for Leu residues.

Figure 1. Ribbon pictures of the structure of HEWL with explicit side chains for which S<sup>(exp)</sup> order-parameter values derived from relaxation measurements are available. Left: Ala, Ile, Leu, Met, Thr and Val side chains; middle: Arg and Trp side chains; right: Asn and Gln side chains.
Table 1. $S_{12}$ values (51) derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures, and the mean of the latter four values and the root-mean-square deviation (RMSD) from it. Order-parameter target values larger than 0.95 were set to 0.95 (second column between brackets). Values differing more than 0.2 from the experimental value (0.95 in case the experimental value is 1) are denoted using italics.

| Residue and methyl group | Experimental value | MD simulation | Mean | RMSD |
|--------------------------|-------------------|---------------|------|------|
| Val2 CG2 | 0.598 | 0.39 | 0.43 | 0.46 | 0.05 |
| Leu8 CD1 | 0.767 | 0.58 | 0.61 | 0.53 | 0.58 | 0.03 |
| Leu8 CD2 | 0.403 | 0.63 | 0.42 | 0.37 | 0.61 | 0.03 |
| Ala9 CB | 1.0 (0.95) | 0.93 | 0.93 | 0.93 | 0.93 | 0.003 |
| Ala10 CB | 0.901 | 0.91 | 0.92 | 0.92 | 0.92 | 0.003 |
| Ala11 CB | 0.861 | 0.91 | 0.92 | 0.92 | 0.92 | 0.001 |
| Met12 CE | 0.812 | 0.33 | 0.26 | 0.58 | 0.41 | 0.12 |
| Leu17 CD1 | 0.630 | 0.46 | 0.33 | 0.52 | 0.48 | 0.10 |
| Leu17 CD2 | 0.632 | 0.49 | 0.37 | 0.55 | 0.50 | 0.08 |
| Leu25 CD1 | 1.0 (0.95) | 0.40 | 0.30 | 0.41 | 0.42 | 0.10 |
| Leu25 CD2 | 0.609 | 0.42 | 0.32 | 0.37 | 0.42 | 0.09 |
| Val29 CG2 | 0.871 | 0.57 | 0.61 | 0.58 | 0.61 | 0.03 |
| Val29 CG2 | 0.791 | 0.37 | 0.60 | 0.57 | 0.60 | 0.03 |
| Ala31 CB | 0.98 (0.95) | 0.94 | 0.93 | 0.93 | 0.94 | 0.004 |
| Thr43 CG2 | 0.361 | 0.67 | 0.79 | 0.62 | 0.72 | 0.07 |
| Thr47 CG2 | 0.327 | 0.73 | 0.67 | 0.68 | 0.70 | 0.02 |
| Thr51 CG2 | 0.778 | 0.49 | 0.58 | 0.64 | 0.46 | 0.54 | 0.07 |
| Ile55 CG2 | 0.739 | 0.49 | 0.61 | 0.71 | 0.63 | 0.09 |
| Ile55 CD | 0.323 | 0.55 | 0.58 | 0.72 | 0.61 | 0.07 |
| Leu56 CD1 | 0.734 | 0.79 | 0.80 | 0.77 | 0.78 | 0.02 |
| Leu56 CD2 | 0.681 | 0.75 | 0.76 | 0.71 | 0.74 | 0.02 |
| ile58 CG2 | 1.0 (0.95) | 0.84 | 0.86 | 0.88 | 0.86 | 0.01 |
| ile58 CD | 0.160 | 0.81 | 0.75 | 0.82 | 0.78 | 0.03 |
| Thr69 CG2 | 0.98 (0.95) | 0.72 | 0.71 | 0.74 | 0.73 | 0.02 |
| Leu75 CD1 | 0.590 | 0.62 | 0.62 | 0.37 | 0.58 | 0.13 |
| Ile78 CG2 | 0.810 | 0.85 | 0.72 | 0.52 | 0.70 | 0.12 |
| ile78 CD | 0.416 | 0.43 | 0.36 | 0.42 | 0.37 | 0.03 |
| Leu83 CD1 | 0.884 | 0.68 | 0.77 | 0.53 | 0.66 | 0.09 |
| Leu83 CD2 | 0.783 | 0.66 | 0.74 | 0.52 | 0.63 | 0.08 |
| Leu84 CD1 | 1.0 (0.95) | 0.46 | 0.66 | 0.63 | 0.60 | 0.08 |
| Leu84 CD2 | 0.879 | 0.45 | 0.64 | 0.60 | 0.58 | 0.08 |
| Ile88 CG2 | 0.697 | 0.55 | 0.62 | 0.80 | 0.70 | 0.11 |
| Ile88 CD | 0.722 | 0.27 | 0.34 | 0.37 | 0.36 | 0.06 |
| Thr89 CG2 | 1.0 (0.95) | 0.71 | 0.66 | 0.72 | 0.68 | 0.03 |
| Ala90 CB | 0.919 | 0.91 | 0.91 | 0.92 | 0.92 | 0.004 |
| Val91 CG1 | 0.764 | 0.63 | 0.75 | 0.76 | 0.76 | 0.09 |
| Val92 CG2 | 0.707 | 0.61 | 0.81 | 0.83 | 0.75 | 0.09 |
| Ala95 CB | 0.680 | 0.94 | 0.94 | 0.94 | 0.94 | 0.01 |
| ile98 CG2 | 0.740 | 0.90 | 0.83 | 0.85 | 0.86 | 0.02 |
| ile98 CD | 0.815 | 0.89 | 0.82 | 0.85 | 0.85 | 0.03 |
| Val99 CG1 | 0.487 | 0.48 | 0.68 | 0.52 | 0.71 | 0.12 |
| Val99 CG2 | 0.517 | 0.85 | 0.80 | 0.53 | 0.71 | 0.12 |
| Met105 CE | 0.630 | 0.80 | 0.36 | 0.39 | 0.52 | 0.18 |
| Ala107 CB | 0.832 | 0.88 | 0.87 | 0.80 | 0.86 | 0.03 |
| Val109 CG2 | 0.354 | 0.36 | 0.38 | 0.51 | 0.41 | 0.06 |
| Val120 CG1 | 0.660 | 0.69 | 0.51 | 0.55 | 0.60 | 0.06 |
| Ala122 CB | 0.879 | 0.78 | 0.67 | 0.82 | 0.80 | 0.03 |
| Ile124 CG2 | 0.753 | 0.75 | 0.66 | 0.72 | 0.68 | 0.07 |
| Ile124 CD | 0.351 | 0.48 | 0.39 | 0.55 | 0.56 | 0.07 |
| Leu129 CD1 | 0.525 | 0.12 | 0.16 | 0.20 | 0.16 | 0.03 |
| Leu129 CD2 | 0.507 | 0.11 | 0.12 | 0.11 | 0.19 | 0.13 | 0.03 |

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 \(^{[11]}\)

$$S_{rot}^2 = \left(3 \cos^2 \beta - 1\right)/2$$

(1)

where $\beta$ 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,

$$S_{CH}^2 = S_{CC}^2 S_{rot}^2$$

(2)

When $\beta = 109.5^\circ$, one has\(^{[11]}\) $S_{CC}^2 = 0.111$. Thus the methyl group restraining is applied to the C–C bond vector and the target value is

$$S_{CC}^2 (exp) = S_{CH}^2 (exp)/0.111$$

(3)
For the NH2 groups in Asn and Gln approximation (2) 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.[22]

### Table 2. *S*2 values (11) for Trp (NE1-HE1) and Arg (NE-HE) side chains derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures, and the mean of the latter four values and the root-mean-square deviation (RMSD) from it. Values differing more than 0.2 from the experimental value are denoted using italics.

| Residue | Experimental value (°) | MD simulation 2VB1 | 4LZT | 1IEE | 1AKI | Mean | RMSD |
|---------|------------------------|-------------------|------|------|------|------|------|
| Trp28   | 0.90                   | 0.88              | 0.84 | 0.85 | 0.87 | 0.86 | 0.02 |
| Trp62   | 0.41                   | 0.73              | 0.66 | 0.75 | 0.57 | 0.68 | 0.07 |
| Trp63   | 0.88                   | 0.83              | 0.81 | 0.85 | 0.78 | 0.82 | 0.03 |
| Trp108  | 0.87                   | 0.87              | 0.80 | 0.89 | 0.61 | 0.79 | 0.11 |
| Trp111  | 0.88                   | 0.83              | 0.78 | 0.78 | 0.80 | 0.80 | 0.02 |
| Trp123  | 0.85                   | 0.70              | 0.68 | 0.61 | 0.66 | 0.66 | 0.03 |
| Arg61   | 0.22                   | 0.22              | 0.30 | 0.33 | 0.32 | 0.29 | 0.04 |
| Arg73   | 0.12                   | 0.24              | 0.17 | 0.40 | 0.19 | 0.25 | 0.09 |
| Arg112  | 0.31                   | 0.28              | 0.16 | 0.23 | 0.18 | 0.21 | 0.05 |
| Arg114  | 0.27                   | 0.13              | 0.32 | 0.19 | 0.19 | 0.21 | 0.07 |
| Arg125  | 0.05                   | 0.12              | 0.14 | 0.10 | 0.14 | 0.13 | 0.02 |

### MD simulations performed

Four unrestrained MD simulations, starting from the four mentioned X-ray crystal structures, were performed:

1. MD_2VB1,
2. MD_4LZT,
3. MD_1IEE,
4. MD_1AKI,

each 20 ns long. The average solute temperatures were 311 K and the solvent temperatures 312 K.

Starting from the 2VB1 X-ray crystal structure, three *S*2-restraining MD simulations were performed:

1. MD_2VB1_Cres, applying *S*2 res restraining,
2. MD_2VB1_Nres, applying *S*2 res restraining,
3. MD_2VB1_C+res, applying *S*2 res and *S*2 Nres restraining,

again each 20 ns long. The average solute temperatures were 311 K and the solvent temperatures 312 K. When restraining an order

### Table 3. *S*2 res values (17) for Asn (ND2-HD21, -HD22) and Gln (NE2-HE21, -HE22) side chains derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures, and the mean of the latter four values and the root-mean-square deviation (RMSD) from it. The assignment in the second column is based on the best agreement with the values of the MD_2VB1 simulation (third column). Values differing more than 0.2 from the experimental value are denoted using italics.

| Residue | Experimental value (°) | MD simulation 2VB1 | 4LZT | 1IEE | 1AKI | Mean | RMSD |
|---------|------------------------|-------------------|------|------|------|------|------|
| Asn19 HD21 | 0.43                 | 0.49              | 0.46 | 0.34 | 0.42 | 0.43 | 0.06 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
| Asn19 HD22 | 0.72                 | 0.82              | 0.60 | 0.62 | 0.70 | 0.69 | 0.09 |
parameter for a bond vector to a target value derived from experiment, the length of the simulation does not play a significant role. It is the restraining force that has to overcome the resistance originating from the particular local protein structure, for example.

Analysis of atomic trajectories

Trajectory energies and atomic coordinates were stored at 5 ps intervals and used for analysis. S' order parameters were calculated using the ensemble averaging expression

\[
S'_{XY} = \frac{1}{2} \left\{ \sum_{i=1}^{3} \sum_{j=1}^{3} \left( \frac{\mu_{XY}(i)\mu_{XY}(j)}{r_{ij}(T)} \right)^2 - \left( \frac{\mu_{XY}(i)}{r_{ij}(T)} \right)^2 \right\} \left( \frac{\rho_{XY}}{r_{ij}} \right)^6
\]

where \( r \) indicates the time-averaging window, here 1 ns, shorter than the rotational correlation time of 5.7 ns of HEWL in solution. \( \mu_{XY}(i) \) and \( \mu_{XY}(j) \) are the components of the vector \( r_{ij} = r_x - r_y \), connecting atoms X and Y, and \( r_{ij} = \sqrt{r_{ij}^2} \) its length. To obtain a dimensionless quantity the term in curly brackets is multiplied with the 6th power of the effective length \( r_{ij}^6 \) of the vector \( r_{ij} \). Because in the present work bond length constraints are used, the length of \( r_{ij} \) is essentially constant over time and thus equal to its effective value \( r_{ij}^6 \).

Before calculating \( S'_{XY} \), the protein trajectory structures are superimposed using the backbone atoms (N, C\(_\alpha\), C) of residues 3–126 in the fit to eliminate the effect of overall rotation of the protein upon the \( S'_{XY} \) values. Use of only the backbone atoms of four of the five \( \alpha \)-helices and two \( \beta \)-strands in HEWL (residues 4–15, 24–36, 41–45, 50–53, 89–99, and 108–115) did not lead to significantly different \( S'_{XY} \) values.

In the \( S' \) order-parameter restraining simulations, the \( S' \) order parameter is calculated at every time step (2 fs) using an exponential damping factor in the average with a memory relaxation time \( r_m = 200 \) ps, and no rotational fit of the protein structures is carried out, which means that the calculated order parameters in the biased simulation are influenced by the stochastic tumbling of the protein. These \( S' \) order-parameter values will thus differ slightly from the ones calculated from the saved trajectory structures, because in the averages in Equation (4) the trajectory structures 5 ps apart are used, the exponential damping factor is not used, the averaging period is 1 ns and a rotational fit of the protein structures is carried out. However, to analyse all trajectories in the same way Equation (4) was used for both the unrestrained and restrained trajectories.

In view of the uncertainty inherent to the derivation of \( S'_{XY}(\text{exp}) \) values from relaxation experiments and inherent to the calculation of \( S'_{XY}(\text{MD}) \) values from MD simulation, a deviation of less than 0.2 between simulation and experiment is considered insignificant.

The GROMOS force fields treat aliphatic carbons as united CH and CH\(_2\) atoms. So inter-hydrogen distances involving the aliphatic hydrogen atoms were calculated using virtual atomic positions for CH and pro-chiral CH\(_2\) and pseudo-atomic positions for CH\(_3\) for those hydrogen atoms. The pseudo-atom NOE distance bound corrections of ref. [44] were used. The set of NOE distance bounds can be found in Table S1 in the Supporting Information, together with values obtained from the seven simulations. The NOE between Trp28 HZ3 and Leu56 HG was reassigned as between Trp28 HZ3 and Leu56 HD\(_4\) following reassessment of the experimental spectra. Inter-hydrogen distances were calculated as \( r^{-3} \), that is, using \( r^{-3} \) averaging over the trajectory structures, where \( r \) indicates the actual hydrogen-hydrogen distance. In view of the uncertainty inherent to the calculation of NOE bounds and \( r^{-3} \) averaged distances, deviations from experiment of less than 0.1 nm are considered insignificant.

For the calculation of the side-chain \( J_{XY} \) couplings, the Karplus relation was used with the parameter values \( a = 9.5 \) Hz, \( b = -1.6 \) Hz and \( c = 1.8 \) Hz. In view of the various factors contributing to an uncertainty of about 2 Hz inherent to the Karplus relation linking structure and \( J \) couplings, a deviation of less than 2 Hz between \( J_{XY} \) coupling values calculated from MD trajectory structures and \( J_{XY} \) coupling derived from experiment is considered insignificant.

Atom-positional root-mean-square differences RMSD(t) between trajectory structures and the X-ray crystal structures and atom-positional root-mean-square fluctuations RMSFs, i.e. around their average positions, in the MD trajectories were calculated after superimposing the backbone atoms (N, C\(_\alpha\), 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.

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 or angle was larger than 135°.

2. Results and Discussion

2.1. Comparison of \( S' \) order-parameter values calculated from unrestrained MD trajectories with NMR derived values

Table 1 lists 51 \( S'_{XY} \) values derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures. The mean of the four MD values and the root-mean-square deviation (RMSD) from it are also presented. As order parameters in an MD simulation can only be equal to 1 if there is no motion at all, order-parameter target values larger than 0.95 were set to 0.95. Deviations from the experimentally derived values of more than 0.2 are denoted in italic. The mean values of the MD simulations show 18 deviations larger than 0.2, 12 are smaller and six are larger than the experimentally derived value. Some large (> 0.8) experimentally derived values, for example, for Met12 CE (0.812), Leu25 CD1 (1.0), Val29 CG1 (0.871), Thr69 CG2 (0.98), Leu84 CD1 (1.0) and CD2 (0.879), Thr89 CG2 (1.0), are not reproduced within 0.2 in any of the four MD simulations. Some small (< 0.4) experimentally derived values, for example, for Thr43 CG2 (0.361), Thr47 CG2 (0.327), Ile55 CD (0.323), and Ile58 CD (0.160), are also not reproduced within 0.2 in any of the four MD simulations. Large order parameters are expected for side chains buried inside the protein, while small order parameters...
are expected for side chains at the protein surface. Yet, this seems not always true. Val99 has in the 2VB1 X-ray structure a solvent accessible area of only 7% and its side chain is surrounded by the side chains of Tyr20, Trp28, Ile98, and Tyr108. This leads to simulated $S_{OH}$ values of about 0.71 to be compared to the experimentally derived value of 0.5. Thr89 has in the 2VB1 X-ray structure a solvent accessible area of 76%. This leads to simulated $S_{OH}$ values of about 0.68 to be compared to the experimentally derived value of 1.0. A larger variation (RMSD $\geq$ 0.12) of $S_{OH}$ values between the four MD simulations is observed for Met12 CE, Leu75 CD1, Ile78 CG2, Val99 CG1 and CG2, and Met105 CE.

Table 2 lists the 11 $S_{OH}$ values for Trp (NE1-HE1) and Arg (NE-HE) side chains derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures. Three simulations show only one deviation from the experimentally derived value larger than 0.2. Both, large and small values are well reproduced.

Table 3 lists the 17 $S_{OH}$ values for Asn (ND2-HD21, -HD22) and Gln (NE2-HE21, -HE22) side chains derived from relaxation measurements and from four unrestrained MD simulations starting from four X-ray crystal structures. The simulations show one or two deviations from the experimentally derived value larger than 0.2, four simulated values, for Asn65 and 74, smaller than the experimentally derived values, and two simulated values, for Asn103 and 113, larger than the experimentally derived ones.

Table 4 shows the occurrence (%) of hydrogen bonds involving the side chains of Arg, Asn, Gln and Trp residues in the four X-ray structures and in the four unrestrained MD simulations starting from the respective four X-ray structures. The four hydrogen bonds present in all four X-ray structures are also observed in the MD simulations, but with widely different occurrences (0–85%). For the hydrogen bonds observed in only three of the four X-ray structures the occurrences vary from 0 to 99%. Considering the different thermodynamic conditions under which the X-ray diffraction of the different crystals was measured and the pH of the NMR measurement in aqueous solution, the observed differences are not surprising.

Figures 2 and S1–S3 show the secondary structure elements of HEWL as a function of time calculated for the four unrestrained MD simulations. Five $\alpha$-helices (red; residues 4–15, 24–36, 89–99, 108–115 and 121–125) and three $\beta$-strands (blue; residues 41–45, 50–53 and 58–59) are largely maintained, but the $\alpha$-helix at residues 108–115 turns into two $\beta$-bridges (yellow) after 3 ns in the MD$_{2VB1}$ simulation (Figure S2). All four simulations show a helix of alternating $\alpha$-helical (red) and $\beta_{\omega}$-helical (black) character at residues 80–85. At residues 21–24, simulation MD$_{1AKI}$ (Figure S3) shows a $\beta_{\omega}$-helix, which is lost after about 6 ns in the MD$_{2VB1}$ (Figure 2) and MD$_{1IEE}$ (Figure S2) simulations, and within 1 ns in the MD$_{4LZT}$ simulation (Figure S1).

Figure 3 shows the backbone C$_\alpha$ atom-positional root-mean-square fluctuations (RMSF) as a function of residue sequence number in the four unrestrained MD simulations MD$_{2VB1}$ (black), MD$_{4LZT}$ (red), MD$_{1IEE}$ (green) and MD$_{1AKI}$ (blue) starting from the respective four X-ray structures. The trajectory structures are translationally and rotationally superimposed using the backbone atoms (N, C$_\alpha$, C) of residues 3–126. The black bars at the top indicate secondary structure elements of HEWL (thick bars: $\alpha$-helix, thin bars, $\beta$-strand).

2.2. Comparison of $S^2$ order-parameter values calculated from $S^2$ order-parameter restraining MD trajectories with NMR derived values

Table 5 lists the 51 $S_{OH}$ values derived from relaxation measurements and from the unrestrained and order-parameter re-
strained MD simulations starting from the 2VB1 X-ray crystal structure using three different sets of $S^2$ order-parameter restraints. The unrestrained MD simulation shows 22 $S^2$ order-parameter values (in italics) that deviate more than 0.2 from the experimentally derived values (in 17 residues: 2, 12, 25, 29, 43, 47, 51, 55, 58, 69, 83, 84, 88, 89, 95, 99 and 129). $S^2$ order-parameter restraining towards the set Cres of 51 target $S^2_{CH}(exp)$ values leads, as expected, to good agreement between simulation and experiment for the 51 $S^2_{CH}$ order parameters. Only three deviations larger than 0.2 are observed, for Ala95 CB and for Leu129 CD1 and CD2. Tables 6 and 7 show that restraining towards the set Cres of 51 target $S^2_{CH}(exp)$ values very

Table 4. Occurrence (%) of hydrogen bonds (52) involving the side chains of Arg, Asn, Gln and Trp residues in four X-ray structures and in the four unrestrained MD simulations starting from the four X-ray structures. Only hydrogen bonds present in one of the X-ray structures or with a population of at least 20% in any of the restrained or unrestrained MD simulations are included. Only hydrogen bond populations of 1% or greater are shown.

| Hydrogen bond | Donor-acceptor | X-ray structure | MD simulation |
|---------------|----------------|----------------|--------------|
|               |                | 2VB1 | 4LZT | IEE | IAKI | 2VB1 | 4LZT | IEE | IAKI |
| Arg5 NE-HE-Trp123 O | 100 100 100 100 | 15 21 10 1 | |
| Arg5 NH1/2-HH2/2-Arg125 O* | 100 100 100 100 | 22 17 24 13 | |
| Asn19 ND2-HD22-Asp18 OD1/2/2 | 11 10 10 10 | 22 17 24 13 | |
| Asn27 ND2-HD22-Trp111 O | 2 2 – – | 2 2 – – | |
| Asn27 ND2-HD22-Ser24 O | 100 100 100 100 | 10 10 10 10 | |
| Asn27 ND2-HD22-Ser24 OG | 100 100 100 100 | 10 10 10 10 | |
| Asn27 ND2-HD22-Cys115 O | – 19 1 1 | – 19 1 1 | |
| Trp28 NE1-He-Leu17 O | – 2 – 29 | – 2 – 29 | |
| Trp28 NE1-He-Tyr23 O | 100 100 100 100 | 32 3 8 – | |
| Gln41 NE2-He-Leu84 O | 100 100 100 100 | 3 2 2 1 | |
| Asn44 ND2-HD22-Asp52 OD1/2/2 | 100 100 100 100 | 100 100 100 100 | |
| Asn44 ND2-HD22-Gln57 OE1 | 100 100 100 100 | 14 2 27 5 | |
| Arg45 NH1-HH12-Gly49 O | 100 100 100 100 | 8 2 – 4 | |
| Asn46 ND2-HD21-Ala107 O | 100 100 100 100 | 56 4 – 4 | |
| Asn46 ND2-HD1/2-Asp52 OD1/2/2 | 100 100 100 100 | 75 40 61 55 | |
| Asn46 ND2-HD22-Ser50 OG | 100 100 100 100 | 1 24 35 10 | |
| Gln57 NE2-He-Glu35 OE1 | 100 100 100 100 | 1 – 36 36 | |
| Gln57 NE2-He-Ala42 O | 100 100 100 100 | 6 7 21 21 | |
| Gln57 NE2-He-Ser36 OG | 100 100 100 100 | 6 46 2 23 | |
| Gln57 NE2-He-Gly54 O | 100 100 100 100 | 85 16 49 49 | |
| Asn59 ND2-HD21-Ser50 OG | 100 100 100 100 | 99 94 79 96 | |
| Asn59 ND2-HD1/2-Asp52 OD1/2/2 | 100 100 100 100 | 70 54 50 50 | |
| Arg61 NH2-HD22-Asp48 OD2 | 100 100 100 100 | 1 – – – | |
| Arg61 NH2-HD21-Asp48 O | 100 100 100 100 | 1 – – – | |
| Arg61 NE-He-Thr69 OG | 100 100 100 100 | 11 – 28 28 | |
| Trp63 NE1-He-Asn106 OD1 | 100 100 100 100 | 52 15 35 35 | |
| Asn65 ND2-HD22-Ans74 OD1 | 100 100 100 100 | 52 15 35 35 | |
| Arg68 NH2-HD2-Thr51 OG1 | 100 100 100 100 | 2 2 2 2 | |
| Arg73 NH1-HH12-Arg61 O | 100 100 100 100 | 13 11 28 28 | |
| Asn74 ND2-HD21-Ans77 O | 100 100 100 100 | 91 59 94 94 | |
| Asn103 ND2-HD22-His98 O | 100 100 100 100 | 42 3 23 23 | |
| Asn106 ND2-HD22-Gly102 O | 100 100 100 100 | 42 3 23 23 | |
| Asn106 ND2-HD22-Ans103 O | 100 100 100 100 | 3 3 3 3 | |
| Asn106 ND2-HD22-Ans103 OD1 | 100 100 100 100 | 3 3 3 3 | |
| Trp108 NE1-He-Leu56 O | 100 100 100 100 | 16 2 1 1 | |
| Trp111 NE1-He-Ans27 OD1 | 100 100 100 100 | 16 2 1 1 | |
| Trp111 NE1-He-Ans27 O | 100 100 100 100 | 16 2 1 1 | |
| Arg112 NE2-Ans106 O | 100 100 100 100 | 20 8 1 1 | |
| Arg112 NH1-HH12-Ans106 O | 100 100 100 100 | 19 28 2 2 | |
| Asn133 ND2-HD21-Val109 O | 100 100 100 100 | 2 – 66 66 | |
| Arg144 NH1-HH12-Glu35 OE1 | 100 100 100 100 | 25 – 25 – | |
| Arg144 NE-He-Ans113 OD1 | 100 100 100 100 | 25 – 25 – | |
| Gln121 NE2-HD22-Asp119 OD1/2/2 | 100 100 100 100 | 22 10 34 34 | |
| Trp123 NE1-He-Gly117 O | 100 100 100 100 | 14 – 20 20 | |
| Trp123 NE1-He-Thr118 OG1 | 100 100 100 100 | 14 8 3 3 | |
| Asn125 NE2-HD22-Asp119 OD2 | 100 100 100 100 | 2 5 1 1 | |
| Asn125 NH2-HD22-Gln121 O | 100 100 100 100 | 1 4 1 1 | |
| Asn125 NH1-HH12-Ala122 O | 100 100 100 100 | 1 4 1 1 | |

* Some hydrogen bonds involving aspartic acid side chains are present in the simulations with either OD1 or OD2 acting as the acceptor. In these cases (marked OD1/2) the highest population of the hydrogen bond involving either OD1 or OD2 is listed. Similarly for hydrogen bonds involving asparagine NH2 groups in some cases (marked ND1/2) the highest population of a hydrogen bond where the donor is either ND2-HD1 or ND2-HD2 is listed while for arginine NH2 groups in some cases (marked NH2-HH2) the highest population of a hydrogen bond where the donor is either NH1-HH11, NH1-HH12, NH2-HH21 or NH2-HH22 is listed.
slightly worsens the agreement between simulation and experiment for the $S_{\alpha 2}$ values (51) derived from relaxation measurements and from the unrestrained and order-parameter restrained MD simulations starting from the 2VB1 X-ray crystal structure. Order-parameter target values larger than 0.95 were set to 0.95 (second column between brackets). Values differing more than 0.2 from the experimental value (0.95 in case the experimental value is 1) are denoted using italics.

Table 5. $S_{\alpha 2}$ values (51) derived from relaxation measurements and from the unrestrained and order-parameter restrained MD simulations starting from the 2VB1 X-ray crystal structure. Order-parameter target values larger than 0.95 were set to 0.95 (second column between brackets). Values differing more than 0.2 from the experimental value (0.95 in case the experimental value is 1) are denoted using italics.

| Residue and methyl group | Experimental value | Unrestrained MD | Order-parameter restrained MD |
|--------------------------|--------------------|-----------------|------------------------------|
|                          |                    | 2VB1 | 2VB1_Cres | 2VB1_Nres | 2VB1_C + Nres |
| Val2 CG2                 | 0.598              | 0.39 | 0.42    | 0.46      | 0.44          |
| Leu8 CD1                 | 0.767              | 0.58 | 0.64    | 0.65      | 0.67          |
| Leu8 CD2                 | 0.803              | 0.63 | 0.72    | 0.72      | 0.71          |
| Ala9 CB                  | 1.0 (0.95)         | 0.93 | 0.93    | 0.93      | 0.93          |
| Ala10 CB                 | 0.901              | 0.91 | 0.92    | 0.92      | 0.92          |
| Ala11 CB                 | 0.861              | 0.91 | 0.92    | 0.92      | 0.92          |
| Met12 CE                 | 0.812              | 0.33 | 0.90    | 0.46      | 0.87          |
| Leu17 CD1                | 0.630              | 0.46 | 0.51    | 0.51      | 0.56          |
| Leu17 CD2                | 0.632              | 0.49 | 0.58    | 0.77      | 0.61          |
| Leu25 CD1                | 1.0 (0.95)         | 0.40 | 0.81    | 0.64      | 0.81          |
| Leu25 CD2                | 0.609              | 0.42 | 0.69    | 0.69      | 0.75          |
| Val29 CG1                | 0.871              | 0.57 | 0.85    | 0.62      | 0.82          |
| Val29 CG2                | 0.791              | 0.57 | 0.84    | 0.61      | 0.81          |
| Ala31 CB                 | 0.98 (0.95)        | 0.94 | 0.95    | 0.93      | 0.94          |
| Thr43 CG2                | 0.361              | 0.68 | 0.43    | 0.74      | 0.35          |
| Thr47 CG2                | 0.327              | 0.73 | 0.40    | 0.73      | 0.42          |
| Thr51 CG2                | 0.778              | 0.49 | 0.85    | 0.54      | 0.79          |
| Ile55 CG2                | 0.739              | 0.49 | 0.63    | 0.53      | 0.69          |
| Ile55 CD                 | 0.323              | 0.55 | 0.41    | 0.56      | 0.40          |
| Leu56 CD1                | 0.734              | 0.79 | 0.63    | 0.56      | 0.74          |
| Leu56 CD2                | 0.681              | 0.75 | 0.66    | 0.53      | 0.68          |
| Ile58 CG2                | 1.0 (0.95)         | 0.84 | 0.87    | 0.72      | 0.88          |
| Ile58 CD                 | 0.160              | 0.81 | 0.20    | 0.78      | 0.22          |
| Thr69 CG2                | 0.98 (0.95)        | 0.72 | 0.89    | 0.84      | 0.88          |
| Leu75 CD1                | 0.590              | 0.62 | 0.69    | 0.52      | 0.56          |
| Ile78 CG2                | 0.810              | 0.85 | 0.83    | 0.59      | 0.80          |
| Ile78 CD                 | 0.416              | 0.43 | 0.45    | 0.29      | 0.40          |
| Leu83 CD1                | 0.884              | 0.68 | 0.79    | 0.35      | 0.77          |
| Leu83 CD2                | 0.783              | 0.66 | 0.72    | 0.41      | 0.71          |
| Leu84 CD1                | 1.0 (0.95)         | 0.46 | 0.86    | 0.60      | 0.86          |
| Leu84 CD2                | 0.879              | 0.45 | 0.83    | 0.58      | 0.84          |
| Ile88 CG2                | 0.697              | 0.55 | 0.66    | 0.66      | 0.66          |
| Ile88 CD                 | 0.722              | 0.27 | 0.64    | 0.49      | 0.58          |
| Thr99 CG2                | 1.0 (0.95)         | 0.71 | 0.87    | 0.69      | 0.85          |
| Ala90 CB                 | 0.919              | 0.91 | 0.92    | 0.92      | 0.91          |
| Val92 CG1                | 0.764              | 0.63 | 0.80    | 0.89      | 0.82          |
| Val92 CG2                | 0.707              | 0.61 | 0.79    | 0.87      | 0.83          |
| Ala95 CB                 | 0.680              | 0.94 | 0.91    | 0.94      | 0.91          |
| Ile98 CG2                | 0.740              | 0.90 | 0.87    | 0.73      | 0.83          |
| Ile98 CD                 | 0.815              | 0.89 | 0.86    | 0.82      | 0.85          |
| Val99 CG1                | 0.487              | 0.85 | 0.38    | 0.53      | 0.45          |
| Val99 CG2                | 0.517              | 0.85 | 0.37    | 0.52      | 0.42          |
| Met105 CE                | 0.630              | 0.80 | 0.76    | 0.37      | 0.79          |
| Ala107 CB                | 0.832              | 0.88 | 0.87    | 0.82      | 0.85          |
| Val109 CG2               | 0.354              | 0.36 | 0.22    | 0.32      | 0.25          |
| Val120 CG1               | 0.660              | 0.69 | 0.52    | 0.52      | 0.57          |
| Ala122 CB                | 0.879              | 0.78 | 0.86    | 0.81      | 0.83          |
| Ile124 CG2               | 0.753              | 0.75 | 0.79    | 0.50      | 0.70          |
| Ile124 CD                | 0.351              | 0.48 | 0.49    | 0.44      | 0.34          |
| Leu129 CD1               | 0.525              | 0.12 | 0.27    | 0.12      | 0.33          |
| Leu129 CD2               | 0.507              | 0.11 | 0.27    | 0.10      | 0.31          |
parameter restraining is not able to reduce the $S_{CH}^2$(MD) value from 0.94 in the unrestrained simulation to the target $S_{CH}^2$(exp) value of 0.68. Enhancing the mobility of the CA–CB vector that is close to the polypeptide backbone and in a residue that is in the centre of a helix seems impossible with the parameters applied here.

Table 5 shows that for the 51 $S_{CH}^2$ order parameters restraining towards the set $Nres$ of 28 target $S_{CH}^2$(exp) values yields 21 deviations larger than 0.2. Restraining towards the $S_{CH}^2$(exp) values does not improve the overall agreement between simulation and experiment for the $S_{CH}^2$ values. Yet, for some $S_{CH}^2$ order parameters the agreement improves (Met12 CE, Leu25 CD1 and CD2, Thr69 CG2, Leu84 CD1 and CD2, Ile88 CD2, Ile98 CG2, Val99 CG1 and CG2) by more than 0.1, and for some $S_{CH}^2$ order parameters the agreement worsens (Leu56 CD1 and CD2, Ile58 CG2, Ile78 CG2 and CD, Leu83 CD1 and CD2, Met105 CE, Val120 CG1 and Ile124 CG2) by more than 0.1.

Table 8 shows the occurrence (%) of hydrogen bonds involving the side chains of Arg, Asn, Gln and Trp residues in the $MD_{2VB1}$ unrestrained MD simulation and in the three $S^2$ order-parameter restraining MD simulations starting from the $2VB1$ X-ray structure. The unrestrained simulation shows nine hydrogen bonds with an occurrence larger than 50%. $S^2$ order-parameter restraining MD simulation reduces this number to 5, 3 and 1 for the three $S^2$ order-parameter restraining MD simulations using the $Cres$, $Nres$ or $C+Nres$ sets of restraints, respectively. All but one of the occurrences of the mentioned 9 hydrogen bonds are reduced by the $S^2$ order-parameter restraining. In contrast, only one hydrogen-bond occurrence is raised above 50% by restraining, the hydrogen bond Asn46 ND2–HD22–Ser50 O, from 2% to 51, 27 and 17%, respectively. Of the 52 hydrogen bonds listed (i.e., observed in either the four X-ray structures or for at least 20% in the seven MD simulations), 38 are observed in the unrestrained simulation, 39, 37 and 42 are observed in the three $S^2$ order-parameter restraining MD simulations using the $Cres$, $Nres$ or $C+Nres$ sets of restraints, respectively.

Table 6. $S_{NH}^2$ values (11) for Trp (NE1–HE1) and Arg (NE–HE) side chains derived from relaxation measurements and from the unrestrained and order-parameter restrained MD simulations starting from the $2VB1$ X-ray crystal structure.

| Residue | Experimental value$^{[44]}$ | Unrestrained MD | Order-parameter restrained MD |
|---------|-----------------------------|------------------|-------------------------------|
|         |                             | $2VB1$           | $2VB1\_Cres$ | $2VB1\_Nres$ | $2VB1\_C+Nres$ |
| Trp28   | 0.90                        | 0.88             | 0.84             | 0.85         | 0.89         |
| Trp62   | 0.41                        | 0.73             | 0.72             | 0.48         | 0.54         |
| Trp63   | 0.88                        | 0.83             | 0.81             | 0.87         | 0.85         |
| Trp108  | 0.87                        | 0.87             | 0.70             | 0.81         | 0.83         |
| Trp111  | 0.88                        | 0.83             | 0.59             | 0.81         | 0.78         |
| Trp123  | 0.85                        | 0.70             | 0.67             | 0.82         | 0.78         |
| Arg61   | 0.28                        | 0.22             | 0.30             | 0.26         | 0.29         |
| Arg73   | 0.12                        | 0.24             | 0.21             | 0.13         | 0.13         |
| Arg112  | 0.31                        | 0.28             | 0.13             | 0.23         | 0.22         |
| Arg114  | 0.27                        | 0.13             | 0.18             | 0.13         | 0.19         |
| Arg125  | 0.05                        | 0.12             | 0.13             | 0.11         | 0.10         |

There are a number of examples where the $MD_{2VB1}$ unrestrained simulation yields a $S^2$ value larger than experiment. When restraining reduces the $S^2$ value, a reduction or change in the populations of hydrogen bonds involving that side chain is observed. For example, Asn44 (experimental $S_{NH}^2$ value 0.51 compared to 0.71 and 0.58 in the $MD_{2VB1}$ and $MD_{2VB1\_C+Nres}$ simulations, respectively, Table 7) shows a reduction in the populations of hydrogen bonds to the side chains of Asp52 and Gln57 (Table 8) and Asn46 (experimental $S_{NH}^2$ value 0.62 compared to 0.82 and 0.56 in the $MD_{2VB1}$ and $MD_{2VB1\_C+Nres}$ simulations, respectively) shows a reduction in the population of the hydrogen bond to the side chain of Asp52, but an increase in the population of hydrogen bonds to both the main chain and side chain oxygens of Ser50 in the restrained simulations. Similarly, for Asn103 (experimental $S_{NH}^2$ value 0.26 compared to 0.61 and 0.17 in the $MD_{2VB1}$ and $MD_{2VB1\_C+Nres}$ simulations, respectively), the hydrogen bond to the side chain of Asp101 is almost completely lost in the restrained simulations.

Figures 4 to 6 show the secondary structure elements$^{[46]}$ of HEWL as a function of time calculated for the three $S^2$ order-parameter restraining MD simulations starting from the $2VB1$ X-ray structure. Compared to the $MD_{2VB1}$ unrestrained simulation (Figure 2), $S_{CH}^2$ order-parameter restraining towards the $S_{CH}^2$(exp) values of set Cres induces some changes in secondary structure as a function of time calculated for the $S_{CH}^2$ order-parameter restraining MD simulation $MD_{2VB1\_Cres}$ starting from the $2VB1$ X-ray structure. Red: $\alpha$-helix; green: $\beta$-helix; black: $3_{10}$–helix; blue: $\beta$-strand; yellow: $\beta$-bridge; brown: bend; grey: turn.
structure (Figure 4). Residues 19–21 become 3_10-helical, the second α-helix (residues 24–36) is slightly more stable at its C-terminal end, the helix of alternating α-helical and 3_10-helical character at residues 80–85 gains 3_10-helical character, residues 103–108 become α-helical after 4 ns, and the α-helix 108–115 gains π-helical character. Generally, the 3_10-helical character is increased. S_α helical restraining towards the S_α(exp) values of set Nres shows similar changes in the secondary structure (Figure 5). The same observation holds for the MD_2VB1_C + Nres simulation (Figure 6).

Figure 7 shows the backbone Cα atom-positional root-mean-square fluctuations (RMSF) as function of residue sequence number in the unrestrained MD simulation MD_2VB1 (black) and in the S_2 order-parameter restraining simulations MD_2VB1_Cres (magenta), MD_2VB1_Nres (cyan) and MD_2VB1_ C + Nres (orange) all starting from the 2VB1 X-ray structure. S_α(CH) order-parameter restraining induces mobility for residues 1–17, residues 85–105 and residues 109–112. Restraining S_α(CH) order parameters shows increased mobility for residues 100–105. Restraining to both sets of order parameters shows increased mobility for residues 100–104 and 109–110.

Table 7. Continued

| Residue | Experimental value | Unrestrained MD | Order-parameter restrained MD |
|---------|--------------------|-----------------|--------------------------------|
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
|        |                    | 2VB1 CRES      | 2VB1 NRES                      |
Table 8. Occurrence (%) of hydrogen bonds (52) involving the side chains of Arg, Asn, Gln and Trp residues in the MD_{2VB1} unrestrained MD simulation and in the three $S^2$ order-parameter restraining MD simulations starting from the 2VB1 X-ray structure. Only hydrogen bonds present in one of the X-ray structures or with a population of at least 20% in any of the restrained or unrestrained MD simulations are included. Only hydrogen bond populations of 1% or greater are shown.

| Hydrogen bond | MD simulation | Order-parameter restrained |
|---------------|---------------|----------------------------|
|               | Unrestrained  | $2VB1_{Cres}$ | $2VB1_{Nres}$ | $2VB1_{C+Nres}$ |
| Arg5 NE-HE-Trp123 O | 15 | 6 | 14 | 9 |
| Arg5 NH1/2-HH1/2-Arg125 O* | 9 | 4 | 7 | 8 |
| Arg5 NH1-HH1-Trp123 O | 22 | 27 | 10 | 12 |
| Asn19 ND2-HD22-Asp18 OD1/2* | 11 | 35 | 30 | 24 |
| Asn27 ND2-HD22-Trp111 O | 2 | 9 | 26 | 4 |
| Asn27 ND2-HD22-Ser24 O | 3 | 3 | 21 | 1 |
| Asn27 ND2-HD22-Cys115 O | 24 | 7 | 6 |
| Trp28 NE1-HE1-Leu17 O | 30 | 30 | 7 |
| Trp28 NE1-HE1-Tyr23 O | 15 | 23 | 4 |
| Gln41 NE2-HE2-Leu84 O | 2 | 2 |
| Asn44 ND2-HD22-Asp52 OD1/2* | 61 | 41 | 30 | 33 |
| Asn44 ND2-HD22-Gln57 OE1 | 14 | 3 | 6 | 3 |
| Arg45 NH1-HH1-Gly49 O | 8 | 2 | 4 | 3 |
| Asn46 ND2-HD21-Ala107 O | 56 | 68 | 41 | 40 |
| Asn46 ND2-HD12-Asp52 OD1/2* | 75 | 25 | 23 | 23 |
| Asn46 ND2-HD22-Ser50 O | 2 | 51 | 27 | 17 |
| Gln57 NE2-HE1-Glu35 OE1 | 1 | – | – | 1 |
| Gln57 NE2-HE1-Ala42 O | 66 | 5 | 24 | 4 |
| Gln57 NE2-HE2-Ser36 OG | 6 | 9 | 4 | 17 |
| Gln57 NE2-HE2-Gly54 O | 85 | 18 | 71 | 8 |
| Asn59 ND2-HD21-Ser50 OG | 99 | 84 | 95 | 93 |
| Asn59 ND2-HD12-Ser52 OD1/2* | 70 | 60 | 52 | 52 |
| Arg61 NH2-HH22-Asp48 OD2 | 1 | – | – | 9 |
| Arg61 NH2-HH21-Asp48 O | – | – | – | – |
| Arg61 NE-HE-Thr69 OG1 | 6 | – | – | – |
| Trp63 NE1-HE1-Asn106 OD1 | 11 | – | – | – |
| Asn65 ND2-HD22-Asn74 OD1 | 52 | 56 | 3 | 2 |
| Arg68 NH2-HH2-Thr51 OG1 | 13 | 2 | 3 |
| Asn68 NH1-HH1-Asp61 O | 13 | 10 | 6 | 11 |
| Asn74 ND2-HD21-Asn77 O | 1 | 1 | 7 | 8 |
| Asn103 ND2-HD22-Ile98 O | – | – | – | – |
| Asn103 ND2-HD22-Asp101 OD1/2 | 42 | 5 | 1 | 4 |
| Asn106 ND2-HD22-Gly102 O | – | 5 | 1 | 1 |
| Asn106 ND2-HD22-Ser50 O | 3 | 3 | – | 1 |
| Asn106 ND2-HD22-Ser52 OD1/3 | 22 | 9 | 24 | 54 |
| Trp111 NE1-HE1-Leu56 O | 91 | 9 | 24 |
| Trp111 NE1-HE1-Asn27 OD1 | 1 | – | 22 | 1 |
| Trp111 NE1-HE1-Asn27 O | – | – | 23 | 19 |
| Arg112 NE-HE1-Asp106 O | 20 | 3 | 4 | 4 |
| Arg112 NH1-HH1-Asn106 O | 19 | 9 | 16 | 6 |
| Asn113 ND2-HD21-Val109 O | 2 | – | 8 | – |
| Arg114 NH1-HH1-Glu35 OE1 | – | 2 | – | 11 |
| Arg114 NE-HE1-Asn113 OD1 | 9 | 2 | – | – |
| Gln121 NE2-HE22-Asp119 OD1/2* | 22 | 12 | 20 | 16 |
| Trp123 NE1-HE1-Gly117 O | 14 | 11 | – | 1 |
| Trp123 NE1-HE1-Thr118 OG1 | 14 | 5 | 45 | 11 |
| Arg125 NE-HE-Gln121 OE1 | 2 | 4 | 5 | 4 |
| Arg125 NH2-HH22-Asp119 OD2 | – | – | – | – |
| Arg125 NH2-HH22-Gln121 OE1 | 1 | 3 | 3 | 3 |
| Arg125 NH1-HH12-Ala122 O | – | – | – | – |

* Some hydrogen bonds involving aspartic acid side chains are present in the simulations with either OD1 or OD2 acting as the acceptor. In these cases (marked OD1/2) the highest population of the hydrogen bond involving either OD1 or OD2 is listed. Similarly for hydrogen bonds involving asparagine NH2 groups in some cases (marked ND2-HD1/2) the highest population of a hydrogen bond where the donor is either ND2-HD21 or ND2-HD22 is listed while for arginine NH2 groups in some cases (marked NH1/2-HH1/2) the highest population of a hydrogen bond where the donor is either NH1-HH11, NH1-HH12, NH2-HH21 or NH2-HH22 is listed.

2.3. Discussion

Table 9 summarises the deviations of $S_{2CH}^C$(MD) values from $S_{2CH}^{exp}$ values for the 51 $S_{2CH}$ order parameters in the seven MD simulations. Of course, $S_{2CH}^C$ order-parameter restraining reduces the number of larger deviations. $S_{2CH}^C$ order-parameter restraining increases the number of deviations larger than 0.1 from 33 in the unrestrained $MD_{2VB1}$ simulation to 35 in the $MD_{2VB1}$...
Figure 6. Secondary structure elements of HEWL as a function of time calculated for the $S_{NH}^2$ and $S_{CH}^2$ order-parameter restraining MD simulations $MD_{2VB1} + Cres$ starting from the $2VB1$ X-ray structure. Red: α-helix; green: π-helix; black: $3_{10}$-helix; blue: β-strand; yellow: β-bridge; brown: bend; grey: turn.

Figure 7. Backbone C atom-positional root-mean-square fluctuations (RMSF) as function of residue sequence number for the unrestrained MD simulation $MD_{2VB1}$ (black) and for the three $S^2$ order-parameter restraining MD simulations $MD_{2VB1}$ Cres (magenta), $MD_{2VB1}$ Nres (cyan) and $MD_{2VB1}$ Cres + Nres (orange) all starting from the $2VB1$ X-ray structure. The trajectory structures are translationally and rotationally superimposed using the backbone atoms (N, C) of residues 3–126. The black bars at the top indicate secondary structure elements of HEWL (thick bars: α-helix; thin bars, β-strand).

$S_{NH}^2$ order-parameter restraining to all 79 experimentally derived $S_{NH}^2$ order-parameter values yields even less deviations larger than 0.1 (12) than restraining only to the 51 $S_{CH}^2$ order-parameter values (16).

Tables 10 and 11 summarise the deviations of $S_{NH}^2$(MD) values from $S_{NH}^2$(exp) values for the 28 $S_{NH}^2$ order parameters in the seven MD simulations. Of course, $S_{CH}^2$ order-parameter restraining marginally improves the agreement with experiment. $S_{CH}^2$ order-parameter restraining does not improve the agreement with experiment for the 28 $S_{NH}^2$ order parameters, from 12 deviations larger than 0.1 in the $MD_{2VB1}$ simulation to 11 deviations in the $MD_{2VB1}$ Nres simulation. Combining $S_{CH}^2$ and $S_{NH}^2$ order-parameter restraining yields almost equally good agreement.

The data in Tables 5 and 7 and 9–11 indicate some unreliability of $S_{CH}^2$ order-parameter restraining, which may have different experimental or computational sources: 1) Some $S_{CH}^2$(exp) values (Table 5) are very different for two vectors in the same side chain, for example, for Ile58 CG2 and CD (1.0 and 0.16) and less so for Leu25 CD1 and CD2 (1.0 and 0.609) and Ile55 CG2 and CD (0.739 and 0.323). This suggests an unlikely large difference in mobility for nearby C–H vectors. 2) Some residues in the protein interior show unexpectedly low $S_{CH}^2$(exp) values, indicating high mobility, for example Val99 (CG1 0.487 and CG2 0.517) with a solvent accessible area in the $2VB1$ X-ray structure of only 7%. 3) Some residues at the surface of the protein show unexpectedly high $S_{CH}^2$(exp) values, indicating low mobility, for example Thr89 CG2 with $S_{CH}^2$(exp) = 1.0 and a solvent accessible area in the $2VB1$ X-ray structure of 76%. 4) As discussed in section 2.3, the $S_{CH}^2$ order-parameter restraining algorithm restrains the C–C vector adjacent to the three C–H vectors of a methyl group, of which the relaxation is measured experimentally. This procedure is based on the assumption that the rotational motion of the C–H vectors around the axis of symmetry of the CH$_3$ group is uniform and decoupled from the motion of the symmetry axis itself (the C–CH$_3$ vector). These assumptions need not be true. This suggests that $S_{CH}^2$ order-parameter restraining is less reliable than $S_{NH}^2$ order-parameter restraining, for which the latter assumptions are not invoked.

Table 12 shows the number of NOE distance bound violations in the four X-ray crystal structures and the seven MD simulations for the 1630 NOE distance bounds specified in Table S1. $S_{CH}^2$ order-parameter restraining decreases the number of NOE bound violations larger than 0.1 nm from 42 in the unrestrained $2VB1$ simulation to 34 in the $MD_{2VB1}$ Cres simulation. $S_{NH}^2$ order-parameter restraining reduces the number

| 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_{2VB1}  | 7                       | 12       | 9       | 6       | 5       | 1       |       |
| MD_{4LZT}  | 6                       | 11       | 8       | 7       | 1       | 1       |       |
| MD_{1IEE}  | 9                       | 13       | 9       | 5       | 1       | 3       |       |
| MD_{1AKI}  | 7                       | 6        | 16      | 8       | 0       | 2       |       |
| MD_{2VB1} Cres | 19                   | 13       | 3       | 0       | 0       | 0       |       |
| MD_{2VB1} Nres | 6                | 11       | 11      | 7       | 2       | 2       |       |
| MD_{2VB1} Cres + Nres | 17           | 11       | 1       | 0       | 0       | 0       |       |
of NOE bound violations larger than 0.1 nm from 42 in the unrestrained MD_2VB1 simulation to 36 in the MD_2VB1_Nres simulation, halving the number of violations larger than 0.2 nm from 13 to 7. Combining simulation, halving the number of violations larger than 0.2 nm.

Table 12. Number of NOE distance bound violations in the four X-ray crystal structures and the seven MD simulations. Number of NOE distance bounds: 1630.

| Structure or simulation | Size of NOE distance bound violation [nm] | 0.05–0.1 | 0.1–0.15 | 0.15–0.2 | 0.2–0.25 | 0.25–0.3 | > 0.3 |
|-------------------------|------------------------------------------|----------|----------|----------|----------|----------|-------|
| 2VB1                    | 22                                      | 7        | 5        | 0        | 0        | 0        | 0     |
| 4LZT                    | 20                                      | 7        | 4        | 0        | 0        | 0        | 0     |
| 1EE                     | 20                                      | 7        | 5        | 0        | 0        | 0        | 0     |
| 1AKI                    | 15                                      | 10       | 4        | 0        | 0        | 0        | 0     |
| MD_2VB1                 | 44                                      | 18       | 11       | 5        | 3        | 5        | 5     |
| MD_4LZT                 | 41                                      | 13       | 13       | 5        | 3        | 5        | 5     |
| MD_1EE                  | 43                                      | 20       | 13       | 8        | 3        | 5        | 5     |
| MD_1AKI                 | 44                                      | 15       | 14       | 2        | 3        | 8        | 8     |
| MD_2VB1_Cres            | 40                                      | 18       | 8        | 4        | 2        | 2        | 2     |
| MD_2VB1_Nres            | 36                                      | 19       | 10       | 4        | 2        | 1        | 1     |
| MD_2VB1_C + Nres        | 42                                      | 14       | 7        | 4        | 4        | 1        | 1     |

Table 13 shows 26 side-chain $3J_{\text{HNH}}$ coupling values for side chains, for which $S^2$ order-parameter values derived from experiment are available, as derived from NMR measurements[23] as well as from the unrestrained and $S^2$ order-parameter restraining MD simulations starting from the 2VB1 X-ray crystal structure. In the unrestrained simulation, six $3J_{\text{HNH}}$ coupling values differ more than 2 Hz from the experimentally derived values, four for side chains for which experimentally derived $S^2_{\text{CH}}$ order-parameter values are available and two for side chains for which experimentally derived $S^2_{\text{NH}}$ order-parameter values are available. $S^2_{\text{CH}}$ order-parameter restraining induces two more deviations of $3J_{\text{HNH}}$-couplings in side chains with $S^2_{\text{CH}}$ order-parameter values, while $S^2_{\text{NH}}$ order-parameter restraining induces four more deviations of $3J_{\text{HNH}}$-couplings in side chains with $S^2_{\text{NH}}$ order-parameter values. Combined $S^2_{\text{CH}}$ and $S^2_{\text{NH}}$ order-parameter restraining also leads to four more deviations of $3J_{\text{HNH}}$-couplings, two in side chains with $S^2_{\text{CH}}$ order-parameter values and two in side chains with $S^2_{\text{CH}}$ order-parameter values. Overall, $S^2$ order-parameter restraining in MD simulation did not improve the agreement with experiment for the 26 side-chain $3J_{\text{HNH}}$ coupling values. For some residues, Thr51 for example, the agreement improves, whereas for other residues, Val29 for example, the agreement worsens.

Overall, the structure of HEWL is maintained in all seven MD simulations, as is indicated in Figures S1 and S2 showing the backbone atom-positional root-mean-square deviation (RMSD) from the 2VB1 X-ray structure as function of time.
3. Conclusions

$S^2$ order parameters for C–H and N–H bonds in proteins derived from NMR relaxation measurements reflect the directional mobility of these bonds. Consequently, they cannot be related to a single protein structure, but to an ensemble of such structures. Therefore, $S^2$ order parameters are not used as data for standard single-structure determination of proteins. A comparison of four X-ray structures of HEWL and four MD simulations starting from these four different X-ray structures illustrates the need of a conformational ensemble representation of the HEWL protein, in particular for its side chains. MD simulation allows for averaging over an ensemble of protein structures, which is used in protein structure determination based on $S^2$ order parameters.\(^\text{103}\) $S^2$ order-parameter restraining can be directly applied to the N–H bond vectors in a protein. In contrast, $S^2_{CH}$ order-parameter restraining of C–H bond vectors in methyl groups makes no sense because of the fast rotation of the three C–H bonds around their symmetry axis parallel to the C–CH$_3$ bond. By assuming the above-mentioned rotation to be uniform and independent from the motion of the symmetry axis itself,\(^\text{104}\) the $S^2_{CH}$ order-parameter restraining algorithm can be applied to the C–CH$_3$ bond. The results for HEWL show that $S^2_{CH}$ order-parameter restraining is more problematic than $S^2_{NH}$ order-parameter restraining, which may be due to less valid assumptions and approximations used to derive experimental $S^2_{CH}$ values from NMR relaxation measurements and the assumptions of uniform rotational motion of methyl C–H bonds around their symmetry axis and of the independence of these motions from each other.

The application of $S^2$ order-parameter restraining to HEWL suggests that this technique is able to produce a conformational ensemble compatible with the experimentally derived $S^2_{CH}$(exp) and $S^2_{NH}$(exp) values. $S^2$ order-parameter restraining in MD simulation does improve the agreement with 1630 NOE atom-atom distance bounds for HEWL. It maintains the overall structure of the protein and induces slightly more mobility, reflected in the backbone atom-positional fluctuations. The unrestrained MD simulations show a high level of conformational disorder for side chains on the protein surface. However, this disorder is increased even further on $S^2$ order-parameter restraining. In the MD simulations, which show good agreement with the experimental order parameters, the populations of many of the hydrogen bonds that are seen in all or most of the X-ray structures are low. This has important implications for the use of X-ray structure data in areas such as drug design, the interpretation of mutational data and receptor binding studies.

Acknowledgement

L.J.S. acknowledges the use of the Advanced Research Computing (ARC) facility at the University of Oxford in carrying out some of the computational work.
Conflict of Interest

The authors declare no conflict of interest.

Keywords: averaging time · conformation sampling · nuclear magnetic resonance · S2 order parameters · structure refinement

[1] W. F. van Gunsteren, J. R. Allison, X. Daura, J. Dolenc, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu, L. J. Smith, Angew. Chem. Int. Ed. 2016, 55, 15990–16010; Angew. Chem. 2016, 128, 16222–16244.

[2] J. Fennem, A. E. Torda, W. F. van Gunsteren, J. Biomol. NMR 1995, 6, 163–170.

[3] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, Chem. Phys. Lett. 1989, 157, 289–294.

[4] P. Groz, W. F. van Gunsteren, W. G. J. Hol, Science 1990, 249, 1149–1152.

[5] C. A. Schiffer, W. F. van Gunsteren, Proteins Struct. Funct. Genet. 1999, 36, 501–511.

[6] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, J. Mol. Biol. 1990, 214, 223–235.

[7] C. A. Schiffer, R. Huber, K. Wüthrich, W. F. van Gunsteren, J. Mol. Biol. 1994, 241, 588–599.

[8] A. E. Torda, R. M. Brunne, T. Huber, H. Kessler, W. F. van Gunsteren, J. Biomol. NMR 1993, 3, 55–66.

[9] N. Hansen, P. Heller, N. Schmid, W. F. van Gunsteren, J. Biomol. NMR 2014, 60, 169–187.

[10] G. Lipari, A. Szabo, J. Am. Chem. Soc. 1982, 104, 4546–4559.

[11] E. R. Henry, A. Szabo, J. Chem. Phys. 1985, 82, 4753–4761.

[12] E. T. Olejniczak, C. M. Dobson, M. Karplus, R. M. Levy, J. Am. Chem. Soc. 1984, 106, 1923–1930.

[13] I. Chandrasekhar, G. M. Clore, A. Szabo, A. M. Gronenborn, B. R. Brooks, J. Mol. Biol. 1992, 226, 239–250.

[14] L. J. Smith, A. E. Mark, C. M. Dobson, W. F. van Gunsteren, Biochemistry 1995, 34, 10918–10931.

[15] D. C. Chatfield, A. Szabo, B. R. Brooks, J. Am. Chem. Soc. 1998, 120, 5301–5311.

[16] V. Kasinath, K. A. Sharp, A. J. Wand, J. Am. Chem. Soc. 2013, 135, 15092–15100.

[17] E. S. O’Brien, A. J. Wand, K. A. Sharp, Prot. Sci. 2016, 25, 1156–1160.

[18] L. J. Smith, W. F. van Gunsteren, N. Hansen, J. Phys. Chem. B 2017, 121, 7055–7063.

[19] L. J. Smith, R. Athish, W. F. van Gunsteren, N. Hansen, Chem. Eur. J. 2017, 23, 9585–9591.

[20] M. Buck, J. Boyd, C. Redfield, D. A. Mackenzie, D. J. Jeenes, D. B. Archer, C. M. Dobson, Biochemistry 1995, 34, 4041–4055.

[21] V. Moorman, K. G. Valentine, A. J. Wand, Prot. Sci. 2012, 21, 1066–1073.

[22] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, J. Mol. Biol. 1993, 229, 930–944.

[23] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, Biochemistry 1991, 30, 986–996.

[24] V. A. Higman, J. Boyd, L. J. Smith, C. Redfield, J. Biomol. NMR 2004, 30, 327–346.

[25] D. Steiner, J. R. Allison, A. P. Eichenberger, W. F. van Gunsteren, J. Biomol. NMR 2012, 53, 223–246.

[26] V. A. Higman, J. Boyd, L. J. Smith, C. Redfield, J. Biomol. NMR 2011, 49, 53–60.

[27] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren, Comp. Phys. Comm. 2012, 183, 890–903.

[28] N. Schmid, J. R. Allison, J. Dolenc, A. P. Eichenberger, A.-P. E. Kunz, W. F. van Gunsteren, J. Biomol. NMR 2011, 51, 265–281.

[29] W. F. van Gunsteren, et al., http://www.gromos.net, GROMOS.

[30] D. Poger, W. F. van Gunsteren, A. E. Mark, J. Comput. Chem. 2010, 31, 1117–1125.

[31] Schmid, A. P. Eichenberger, A. Choutko, S. Rinker, M. Winger, A. E. Mark, W. F. van Gunsteren, Eur. Biophys. J. 2011, 40, 843–856.

[32] K. Bartik, C. Redfield, C. M. Dobson, Biophys. J. 1994, 66, 1180–1184.

[33] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in Intermolecular Forces (Ed.: B. Pullman), Reidel, Dordrecht, 1981, pp. 331–342.

[34] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, J. Comput. Phys. 1977, 23, 327–341.

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

[36] W. F. van Gunsteren, H. J. C. Berendsen, R. G. Geurtsen, H. R. J. Zwijer- man, Ann. New York Acad. Sci. 1986, 482, 287–303.

[37] J. A. Barker, R. O. Watts, Mol. Phys. 1973, 26, 789–792.

[38] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, J. Chem. Phys. 1985, 82, 5451–5459.

[39] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, J. Chem. Phys. 2001, 115, 1125–1136.

[40] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, Nucleic Acids Res. 2000, 28, 235–242.

[41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, J. Chem. Phys. 1984, 81, 3684–3690.

[42] 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, J. Chem. Theory Comput. 2011, 7, 3379–3390.

[43] W. F. van Gunsteren, R. Boelens, R. Kappetein, R. M. Scheek, E. R. P. Zuideweg in Molecular Dynamics and Protein Structure (Ed.: J. Hermans), Polycrystal Book Service, Western Springs, Illinois, 1985, pp 92–99.

[44] K. Wüthrich, M. Billette, W. Braun, J. Mol. Biol. 1983, 169, 949–961.

[45] A. E. Torda, W. F. van Gunsteren, Comp. Phys. Commun. 1991, 62, 289–296.

[46] M. Karplus, J. Chem. Phys. 1959, 30, 11–15.

[47] M. Karplus, J. Amer. Chem. Soc. 1963, 85, 2870–2871.

[48] A. deMarco, M. Lilnáz, K. Wüthrich, Biopolymers 1978, 7, 617–636.

[49] W. Kabsch, C. Sander, Biopolymers 1983, 22, 2577–2637.