Microscopic analysis of protein oxidative damage: effect of carboxylation on structure, dynamics and aggregability of villin headpiece

*Drazen Petrov*¹,²,³ and *Bojan Zagrovic*¹,²,³,*

*Supplementary Information*

*(revised manuscript)*

¹Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna, AT-1030, Austria

²Mediterranean Institute for Life Sciences, Mestrovicevo setaliste bb, Split, HR-21000, Croatia

³Department of Physics, Faculty of Science, University of Split, Teslina 12, Split, HR-21000, Croatia

*Corresponding author e-mail: bojan.zagrovic@univie.ac.at
Parameterization of Asa and Gsa residues

Lysine residues were changed into Asa by removing the hydrogen atoms (HZ1, HZ2 and HZ3) bonded to the nitrogen atom NZ in the side chain, and by replacing the nitrogen atom (NZ) with oxygen; arginine residue was changed into Gsa by removing all side chain atoms, except the three carbon atoms (CB, CG and CD) and the nitrogen atom NE, which was replaced with an oxygen atom; finally the proline residue was changed into Gsa the same way like arginine by replacing the hydrogen atom HD2 bound to CD carbon atom with oxygen. The double bond between carbon and oxygen atoms, and the potential energy term for an angle between a triplet of one oxygen and two carbon atoms in the GROMOS 45A3 force field\(^1\) are described always using the same parameters regardless of the type of carbon or oxygen atoms (e.g. the carboxyl group in the backbone). In order to be internally consistent, we used the same bond parameters in the description of the bond between the atoms CD and oxygen in Asa and the atoms CG and oxygen in Gsa, and the same angle parameters in the description of the angle between the last two carbon atoms and oxygen atom both in Asa and Gsa. All other bonded parameters in Asa and Gsa residues were the same as in lysine and arginine residues, respectively. GROMOS 45A3 building block files for Asa and Gsa are given at the end of the Supporting Information.

Thermodynamic integration:

We used thermodynamic integration (TI)\(^2\) for the calculation of hydration free energies for all amino-acid side-chain analogues and two carbonylated amino-acid side-chain analogues in order to estimate the difference in hydration free-energy between Lys and Asa; and Arg and Gsa. Hydration free energies of neutral and charged forms of native and carbonylated amino-acid side-chain analogues with the CB atom CH\(_n\) replaced by CH\(_{n+1}\), were calculated using TI. Since only hydration free energies of neutral forms of amino-acid side-chain analogues are experimentally measurable, we used them for direct comparison between calculated and experimental values.
The equilibration and free energy calculations in water were carried out using the same conditions as described above for protein simulations, while calculations in vacuo were carried out at 300 K (Berendsen thermostat and $\tau_T = 0.05$ ps were used) without periodical boundary condition, using the simple cut-off method for calculating electrostatics with a cutoff of $r_c=1.4$ nm. Non-bonded interactions of side chains were scaled down to zero in a stepwise manner using a coupling parameter $\lambda$. Free energy changes upon removal of non-bonded interactions were calculated as integrals of the averages of the derivatives of the total system Hamiltonian with respect to $\lambda$, between the boundaries $\lambda=0$ and $\lambda=1$,

$$\Delta G = \int_0^1 \left( \frac{\partial H}{\partial \lambda} \right) d\lambda.$$  

After initially 26 evenly spaced $\lambda$-points were sampled, changes in slope at each point of the $\partial H / \partial \lambda$ versus $\lambda$ graph were calculated. The number of additional $\lambda$-points placed between two given neighbor $\lambda$-points in the second step was proportional to the sum of slope changes at these points, for a total of 26 additional $\lambda$-points. The slope changes for the first and the last $\lambda$-point were considered to be the same as the slope changes in the second and the penultimate $\lambda$-point respectively. Trapezoidal integration was used to evaluate the integral in equation (1) using 52 $\lambda$-points. Sampling of 50 ps of equilibration and 200 ps of data collection at each point were used. In order to avoid singularities in the non-bonded interaction a soft-core interaction was used

$$V_{SC} (r) = (1 - \lambda)V^A (r_A) + \lambda V^B (r_B),$$  

$$r_A = (\alpha \sigma_A^6 \lambda^p + r^6)^{1/6},$$  

$$r_B = (\alpha \sigma_B^6 (1 - \lambda^p) + r^6)^{1/6},$$

where $\sigma_A$ and $\sigma_B$ are van der Walls parameters and $\alpha=1.51$ and $p=1.3^4$ The hydration free energy was calculated by subtracting the free energy change when side chain was simulated in vacuo from the free energy change when side chain was simulated in water.
Fractional contribution of different specific factors to villin headpiece destabilization upon carbonylation

To find how much different specific factors contribute to destabilization of villin headpiece, using our simulated results, we have derived a set of 20 inequalities.

Inequalities based on RMSD:

\[ 5\Delta G_K < \Delta \Delta G_{f-u}, \] \hspace{1cm} (5)
\[ \Delta G_K + \Delta G_R < \Delta \Delta G_{f-u}, \] \hspace{1cm} (6)
\[ 2\Delta G_K + \Delta G_R > \Delta \Delta G_{f-u}, \] \hspace{1cm} (7)
\[ 2\Delta G_K + \Delta G_p < \Delta \Delta G_{f-u}, \] \hspace{1cm} (8)
\[ 3\Delta G_K + \Delta G_p > \Delta \Delta G_{f-u}, \] \hspace{1cm} (9)
\[ \Delta G_R + \Delta G_p < \Delta \Delta G_{f-u}, \] \hspace{1cm} (10)
\[ \Delta G_K + \Delta G_R + \Delta G_p > \Delta \Delta G_{f-u}. \] \hspace{1cm} (11)

Inequalities based on core compactness:

\[ 5\Delta G_K < \Delta \Delta G_{f-u}, \] \hspace{1cm} (12)
\[ 3\Delta G_K + \Delta G_R < \Delta \Delta G_{f-u}, \] \hspace{1cm} (13)
\[ 4\Delta G_K + \Delta G_R > \Delta \Delta G_{f-u}, \] \hspace{1cm} (14)
\[ 5\Delta G_K + \Delta G_p < \Delta \Delta G_{f-u}, \] \hspace{1cm} (15)
\[ 2\Delta G_K + \Delta G_R + \Delta G_p < \Delta \Delta G_{f-u}, \] \hspace{1cm} (16)
\[ 3\Delta G_K + \Delta G_R + \Delta G_p > \Delta \Delta G_{f-u}. \] \hspace{1cm} (17)

Inequalities based on core α-helicity:

\[ 5\Delta G_K < \Delta \Delta G_{f-u}, \] \hspace{1cm} (18)
\[ 3\Delta G_K + \Delta G_R < \Delta \Delta G_{f-u}, \] \hspace{1cm} (19)
\[ 4\Delta G_K + \Delta G_R > \Delta \Delta G_{f-u}, \] \hspace{1cm} (20)
$$3\Delta G_K + \Delta G_p < \Delta \Delta G_{f-u},$$  
$$4\Delta G_K + \Delta G_p > \Delta \Delta G_{f-u},$$  
$$2\Delta G_K + \Delta G_R + \Delta G_p < \Delta \Delta G_{f-u},$$  
$$3\Delta G_K + \Delta G_R + \Delta G_p > \Delta \Delta G_{f-u}.$$  

To find a solution to this set of inequalities numerically, we discretized the space of values for $\Delta \Delta G_K$, $\Delta \Delta G_R$ and $\Delta \Delta G_P$ to integer percentage of $\Delta G_{f-u}$ and counted the number of fulfilled inequalities for a given set of discrete values of $\Delta \Delta G_K$, $\Delta \Delta G_R$ and $\Delta \Delta G_P$, among all 1030301 variations for $0 \leq \Delta \Delta G_K \leq 100$, $0 \leq \Delta \Delta G_R \leq 100$ and $0 \leq \Delta \Delta G_P \leq 100$ in steps of $0.01\Delta G_{f-u}$. The number of fulfilled inequalities ranged from 6 to 16.
Asa and Gsa GROMOS 45A3 parameters

ffG45a3.rtp file:

[ ASA ]
[ atoms ]
  N   N   -0.28000   0
  H   H    0.28000   0
  CA  CH1    0.00000   1
  CB  CH2    0.00000   1
  CG  CH2    0.00000   1
  CD  CH2    0.00000   1
  CE  CH1    0.38000   2
  OE1    O    -0.38000   2
  C     C     0.380     3
  O     O     -0.380     3
[ bonds ]
  N   H    gb_2
  N   CA    gb_20
  CA  C    gb_26
  C   O    gb_4
  C   +N    gb_9
  CA  CB    gb_26
  CB  CG    gb_26
  CG  CD    gb_26
  CD  CE    gb_26
  CE  OE1    gb_4
[ angles ]
;  ai   aj   ak   gromos type
  -C   N   H   ga_31
  H   N   CA   ga_17
  -C   N   CA   ga_30
  N   CA   C   ga_12
  CA  C   +N   ga_18
  CA  C   O   ga_29
  O   C   +N   ga_32
  N   CA  CB   ga_12
  C   CA  CB   ga_12
  CA  CB  CG   ga_14
  CB  CG  CD   ga_14
  CG  CD  CE   ga_14
  CD  CE  OE1   ga_29
[ impropers ]
;  ai   aj   ak   al   gromos type
  N   -C   CA   H   gi_1
  C   CA   +N   O   gi_1
  CA  N   C   CB   gi_2
[ dihedrals ]
;  ai   aj   ak   al   gromos type
  -CA  -C   N   CA   gd_4
  -C   N   CA   C   gd_19
  N   CA   C   +N   gd_20
N  CA  CB  CG   gd_17
CA  CB  CG  CD   gd_17
CB  CG  CD  CE   gd_17
CG  CD  CE  OE1  gd_20

[ GSA ]
[ atoms ]
N    N   -0.28000     0
H    H    0.28000     0
CA   CH1   0.00000     1
CB   CH2   0.00000     1
CG   CH2   0.00000     1
CD   CH1   0.38000     2
OE1    O  -0.38000     2
C    C     0.380     3
O    O   -0.380     3

[ bonds ]
N    H  gb_2
N   CA  gb_20
CA   C  gb_26
C    O  gb_4
C   +N  gb_9
CA   CB  gb_26
CB   CG  gb_26
CG   CD  gb_26
CD  OE1  gb_4

[ angles ]
; ai   aj   ak   gromos type
-CA    -C   N     H   ga_31
  H   N    CA   ga_17
-CA    -C   N    CA   ga_30
  N   CA    C   ga_12
CA   C    +N   ga_18
CA   C    O   ga_29
  O   C    +N   ga_32
N    CA   CB   ga_12
  C    CA   CB   ga_12
CA   CB   CG   ga_14
CB   CG   CD   ga_14
CG   CD  OE1  ga_29

[ impropers ]
; ai   aj   ak   al   gromos type
  N    -C   CA   H    gi_1
  C    CA   +N   O    gi_1
  CA    N   C    CB    gi_2

[ dihedrals ]
; ai   aj   ak   al   gromos type
-C    -CA    -C   N    CA   gd_4
-C    N     CA    C    gd_19
N    CA    C    +N   gd_20
N    CA   CB    CG   gd_17
CA   CB   CG   CD   gd_17
CB   CG   CD  OE1  gd_20
ffG45a3.hdb file:

ASA 1
  1 1 H N -C CA

GSA 1
  1 1 H N -C CA
**Figure Captions**

**Figure S1.** Hydration free energies of native and carbonylated side-chains analogues. Comparison of experimental and calculated hydration free energies using thermodynamic integration. Large black diamonds represent carbonylable sidechains Lys, Arg in neutral forms and their carbonylated sidechains, aminoacidic semialdehyde (Asa) and glutamic semialdehyde (Gsa), while small gray diamonds represent all other sidechains. The regression line with $R^2=0.804$ shows that experimental and calculated values are well correlated. Inset - differences ($\Delta\Delta G_{\text{hydr}}$) between hydration free energies of Asa and Lys; and Gsa and Arg from experiment and simulation. Note that the difference in hydrophobicity is significantly greater than estimated by thermodynamic integration simulation because lysine and arginine residues at biologically relevant pH are usually in charged forms, which are much more hydrophilic than their neutral forms used here.

**Figure S2.** Two dimensional projection of the free-energy landscape as a function of MHP and RMSD from the experimental NMR structure for the last 25 ns of all the simulations. Since ensembles with different number of carbonylated residues contain structures form different number of simulations (e.g. native ensemble contains structures from 5 simulations while 5-carbonylations ensemble contains structures from 21 simulations), the fractions of total population are rescaled in such a way so that the total sums of the fractions for each ensemble are equal. The relative free energy values were calculated as the negative logarithm of the rescaled fractions. The ellipses are centered at the average values of MHP and RMSD with the major and minor semi-axes equal to the standard deviations of the distributions.
**Figure S3.** Site-specific intrinsic aggregation propensity, Zagg, for: (a) The native (solid black line), fully carbonylated (dashed black line) villin headpiece and averages over all combinations of villin headpiece with 1, 2, 3, 4, 5 and 6 carbonylated residues (thin solid gray lines); comparison between villin headpiece in the native state and with one carbonylated residue: (b) native - solid black line, Lys8 to Asa - dashed gray line; (c) native - solid black line, Lys25 to Asa - dashed gray line; (d) native - solid black line, Lys30 to Asa - dashed gray line; and (e) native - solid black line, Lys33 to Asa - dashed gray line.

**Table S1.** Comparison of basic physico-chemical properties of Asa, Gsa, Leu and Val. Solvent accessible surface area (SASA) and molecular volume were calculated using Gromacs tools.\(^5\)
References

(1) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *Journal of Computational Chemistry* **2001**, *22*, 1205.

(2) Beveridge, D. L.; DiCapua, F. M. *Annual review of biophysics and biophysical chemistry* **1989**, *18*, 431.

(3) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529.

(4) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C.; [www.gromacs.org](http://www.gromacs.org); 2005.

(5) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
Figure S1.

The graph shows a linear regression line with the equation $y = 0.7485x + 5.3619$ and an $R^2$ value of 0.8041. The graph plots the calculated $\Delta G_{\text{hyd}}$ (kJ mol$^{-1}$) against the experimental $\Delta G_{\text{hyd}}$ (kJ mol$^{-1}$). The points for Lys $\rightarrow$ Asa, Asa, Gsa, Lys, and Arg are marked with different symbols. The inset shows a histogram for $\Delta G_{\text{hyd}}$ values for charged and neutral states compared to experimental data.
Figure S2
Figure S3.
Table S1.

| Properties          | Asa   | Leu   | Gsa   | Val   |
|---------------------|-------|-------|-------|-------|
| Chemical formula    | C₅H₆O | C₅H₁₀ | C₃H₆O | C₃H₈  |
| MHP                 | 1.56  | 1.67  | 1.27  | 1.28  |
| Mw (g mol⁻¹)        | 72.11 | 58.12 | 58.08 | 44.1  |
| SASA (nm²)          | 1.9   | 1.8   | 1.7   | 1.6   |
| Volume (nm³)        | 0.23  | 0.21  | 0.19  | 0.18  |
| Charge              | 0     | 0     | 0     | 0     |