



JOURNAL OF  
SYNCHROTRON  
RADIATION

**Volume 22 (2015)**

**Supporting information for article:**

**Identifying and quantifying radiation damage at the atomic level**

**Markus Gerstel, Charlotte M. Deane and Elspeth F. Garman**

# Identifying and quantifying radiation damage at the atomic level

Gerstel, M., Deane, C. M. & Garman, E. F. (2015).  
 J. Synchrotron Rad. **22**, doi:10.1107/S1600577515002131.

**Table S1** Isotropic atomic *B*-factors present in the PDB subset.

Atomic <i>B</i> -factor [ $\text{\AA}^2$ ]	Number of atoms	Percentage of atoms
$B < 0$	3	<0.01 %
$B = 0$	4,610	0.05 %
$0 < B < 80$	10,143,262	99.65 %
$B \geq 80$	31,338	0.31 %

Atoms with negative atomic *B*-factors were found in PDB structure entry 3cls.

**Table S2** Occupancy values present in the PDB subset.

Occupancy <i>occ</i>	Number of atoms	Percentage of atoms
$occ < 0$	5	<0.01 %
$occ = 0$	9,177	0.09 %
$0 < occ < 0.5$	68,511	0.67 %
$occ = 0.5$	234,192	2.30 %
$0.5 < occ < 1$	72,020	0.71 %
$occ = 1$	9,795,303	96.23 %
$occ > 1$	5	<0.01 %

Some PDB structures contain occupancy values outside the valid probability range: the structure entries 1ie0 and 1wul report occupancies above 1, and the structure 3hkw reports occupancies below 0.

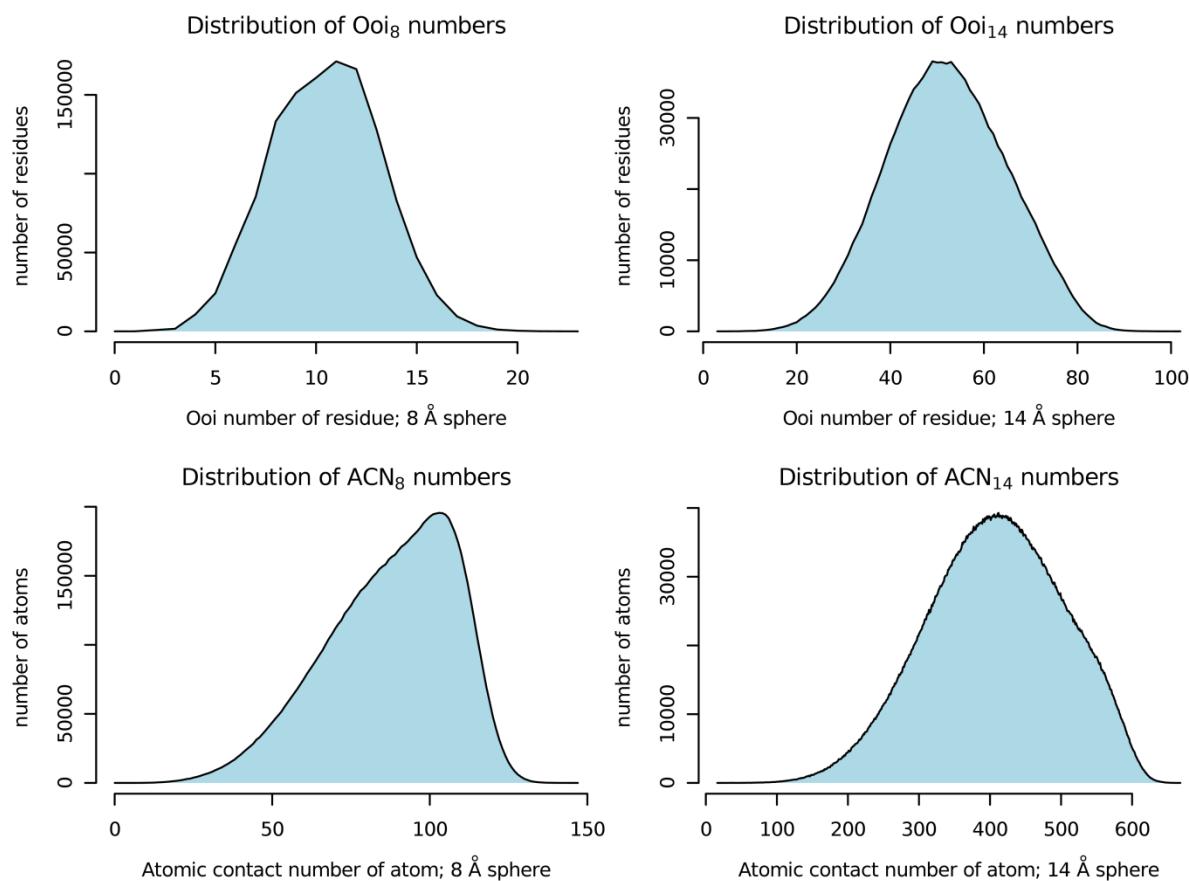
**Table S3** Packing densities of protein secondary structure.

Density metric, radius	Secondary structure	Min	Mean	Max	Variance	Secondary structure	Min	Mean	Max	Variance
Ooi, 8 Å	b	4	11.5	21	6.1	H	4	10.5	22	5.5
	C	1	9.7	22	8.5	I	4	12.4	20	6.9
	E	4	12.0	23	5.1	T	8	9.6	22	8.1
	G	4	9.5	21	7.1					
Ooi, 13 Å	b	10	43.4	72	89.5	H	8	42.9	80	103.6
	C	3	39.7	78	106.2	I	22	48.8	71	136.1
	E	11	47.0	84	93.6	T	5	38.6	82	106.8
	G	8	39.6	75	104.2					
Ooi, 14 Å	b	11	53.1	90	132.1	H	8	52.6	100	148.6
	C	3	48.9	95	153.9	I	29	59.9	82	211.8
	E	12	58.0	102	141.3	T	5	47.8	96	154.3
	G	9	49.1	96	150.4					
ACN, 8 Å	b	13	92.3	141	333.9	H	8	89.2	147	360.8
	C	0	81.1	145	444.8	I	22	98.8	141	337.2
	E	8	96.5	145	294.2	T	8	79.9	144	447.6
	G	8	82.2	141	424.6					
ACN, 13 Å	b	72	340.6	531	5,245.9	H	46	336.2	542	5,757.5
	C	11	313.7	541	6,032.7	I	158	385.7	517	7,705.1
	E	46	366.8	547	5,191.2	T	35	306.7	540	6,229.6
	G	48	314.5	539	6,074.3					
ACN, 13.5 Å	b	74	376.7	589	6,416.0	H	50	372.4	600	6,973.5
	C	14	348.1	597	7,293.6	I	173	428.4	570	9,721.7
	E	53	405.5	601	6,377.9	T	37	340.8	597	7,541.5
	G	51	349.3	602	7,341.6					
ACN, 14 Å	b	80	415.3	654	7,790.0	H	52	410.9	666	8,366.2
	C	16	385.1	658	8,757.2	I	190	474.0	633	12,043.5
	E	56	446.8	668	7,762.1	T	38	377.4	666	9,056.6
	G	57	386.4	659	8,809.8					

Range, mean and variance of different packing density metrics applied to the 2,704 selected PDB structures.

Both ACN and Ooi metrics are counts, so minimum and maximum values are integers. Mean and variance are given to 1 dp. The seven secondary structure labels are: alpha helix (H; n=3,323,956), 3-10 helix (G; n=428,747),  $\pi$ -helix (I; n=1,515), extended conformation (E; n=2,351,831), isolated bridge (b; n=120,856), turn (T; n=1,859,115) and coil (C; n=1,679,478) (Frishman & Argos, 1995).





**Figure S1** Distribution of Ooi numbers and atomic contact numbers (ACN) for radii of 8 Å and 14 Å for the 2,704 selected PDB structures. Note that both Ooi numbers and ACNs are discrete metrics; they can only take integer values.

1. Let the variable  $a$  denote a single atom.
2. Let the set  $A$  describe the set of all non-hydrogen protein atoms within the same PDB structure.
3. Choose a packing density function  $P: A \rightarrow \mathbb{R}$  mapping every atom to a packing density. For this paper  $P$  was defined using atomic contact numbers with a range of 14 Å.
4. Let  $B: A \rightarrow \mathbb{R}$  map every atom to its atomic B factor (atomic displacement parameter).
5. The function of average B factors returns the arithmetic mean of the atomic B factors of a set of atoms  $X \subseteq A$ :

$$\bar{B}(X) := \frac{\sum_{a \in X} B(a)}{|X|}$$

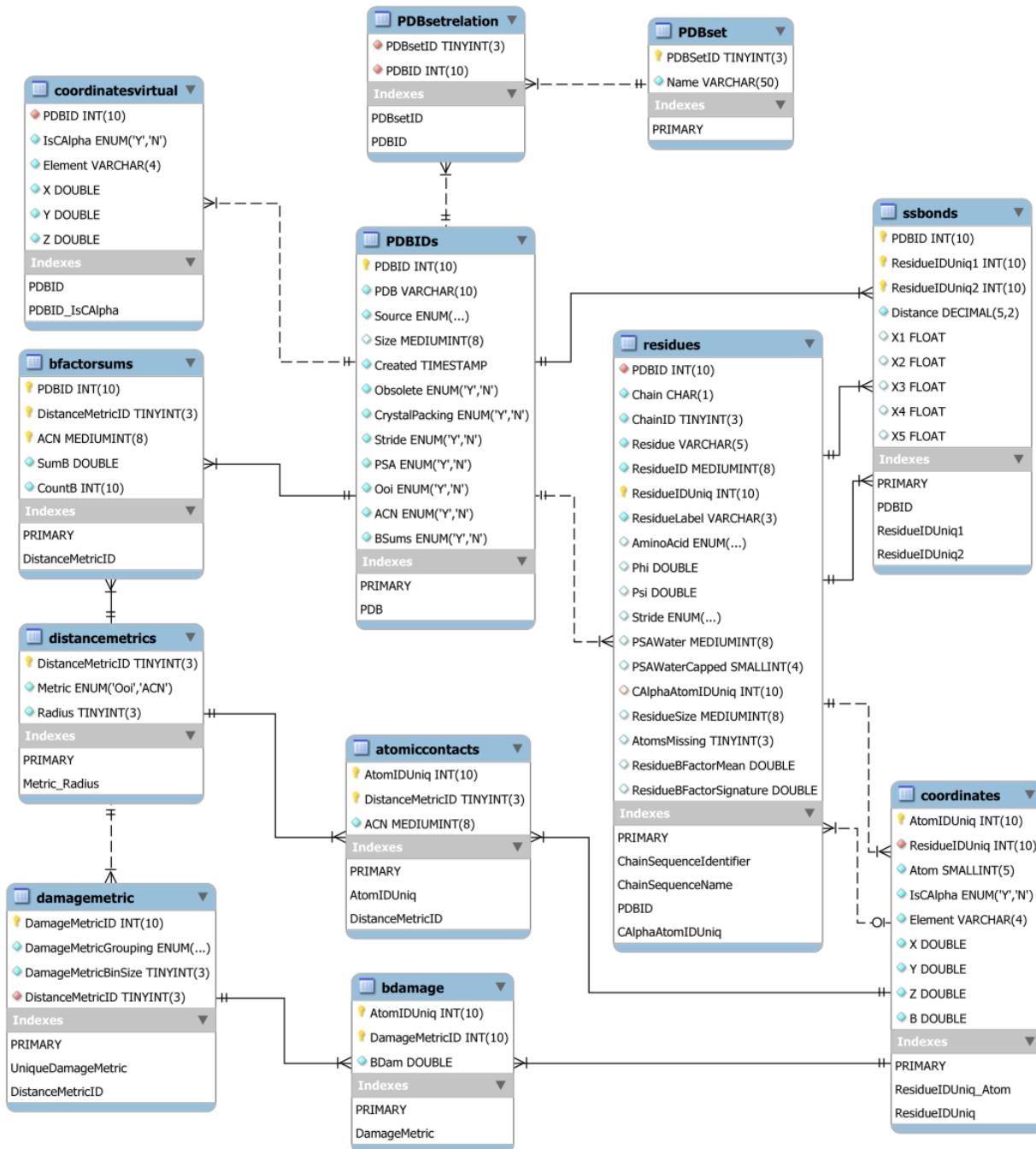
6. Define the packing density environments  $E_n := \{a \in A \mid P(a) = n\}$  as sets, that contain all atoms with the packing density  $n$  as determined by  $P$ . The set of all nonempty  $E_n$  is a partition of  $A$ .
7. Similar packing density environments  $S_m$  are defined as sets of all atoms with a packing density *similar* to  $m \in \mathbb{R}$ :

$$S_m := \bigcup_{n \in sim(m)} E_n$$

8. Similarity can be defined by range, e.g.  $sim(m) := \{x \in \mathbb{R} \mid m - \varepsilon \leq x \leq m + \varepsilon\}$ ,  $\varepsilon \geq 0$  or by binning, e.g.  $sim(m) := \left\{x \in \mathbb{R} \mid \left\lfloor \frac{x}{z} \right\rfloor = \left\lfloor \frac{m}{z} \right\rfloor\right\}$  with a bin size  $z > 0$ . Here the latter definition was used with  $z = 10$ .
9. Define  $B_{\text{Damage}}$  of an atom  $a$  as the atomic B factor divided by the average B factor of all non-hydrogen protein atoms within the same PDB structure and with a similar packing density as  $a$ :

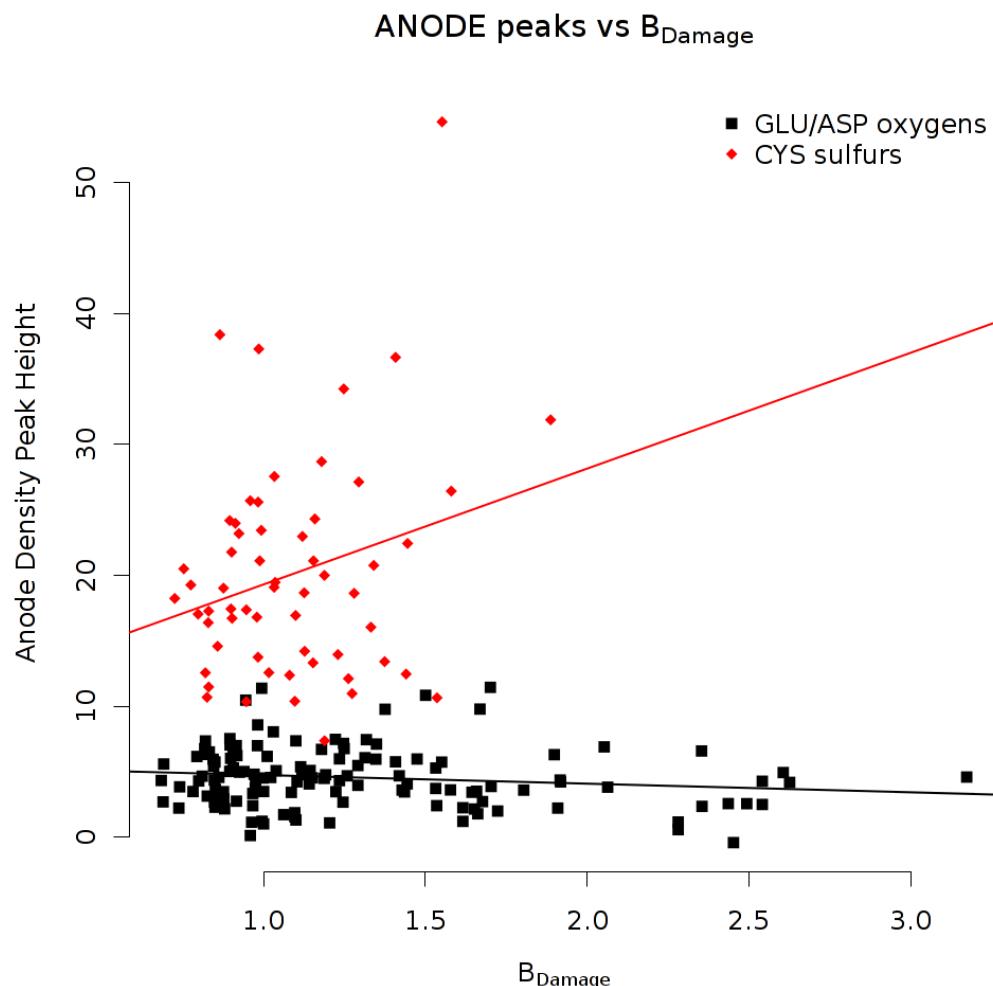
$$B_{\text{Damage}}(a) := \frac{B(a)}{\bar{B}(S_{P(a)})}$$

**Figure S2** Mathematical description of  $B_{\text{Damage}}$  calculation.



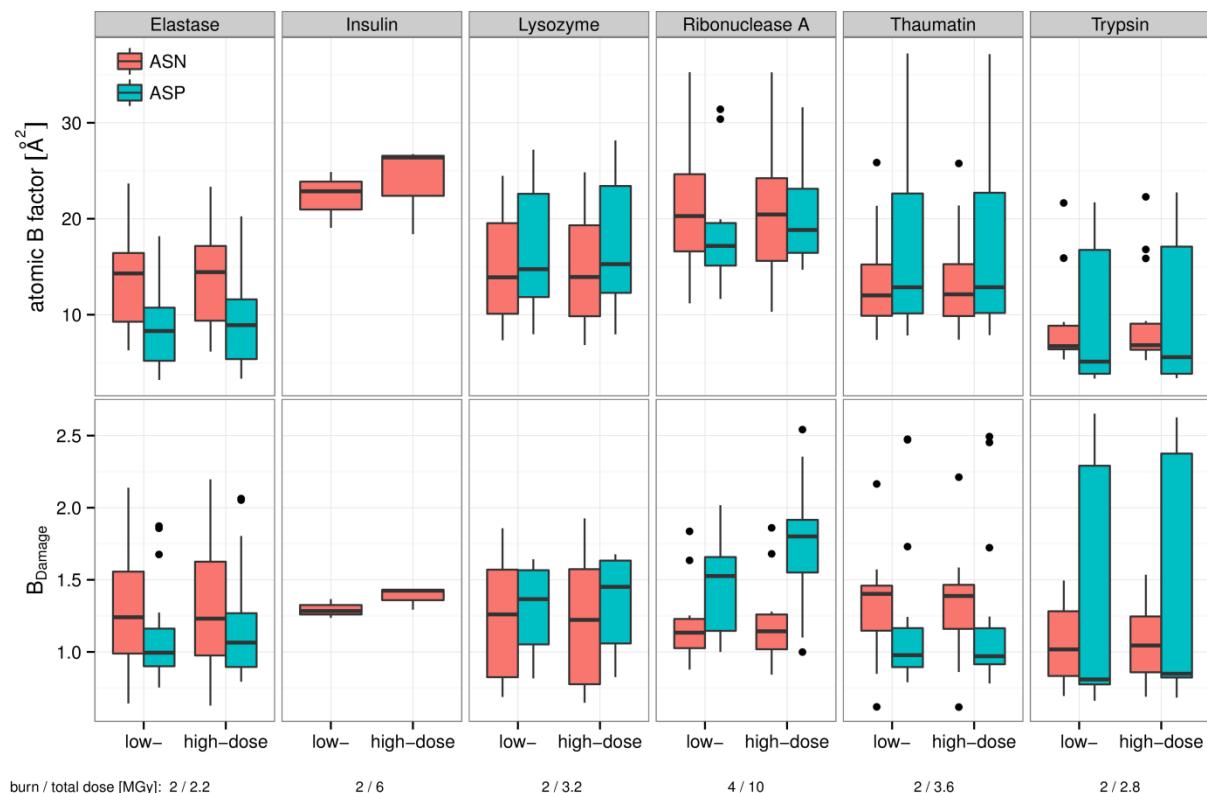
**Figure S3** Entity-Relationship diagram of the relational database used for this investigation.

Diagram prepared using MySQL Workbench 6.0 (<http://mysqlworkbench.org/>).



**Figure S4** Peak heights of electron density difference maps around the cysteine sulfurs (red) and the terminal oxygens of glutamic and aspartic acids (black) plotted against their  $B_{\text{Damage}}$  values for the six Nanao protein datasets. Lines show least squares fits to the points. Peak heights were calculated using SHELXC (Sheldrick, 2010) and ANODE (Thorn & Sheldrick, 2011).  $B_{\text{Damage}}$  values were calculated for the relevant high-dose dataset.

There is some evidence for a correlation of high  $B_{\text{Damage}}$  values and strong electron density peak differences for cysteine sulfurs ( $\text{CC}=0.25$ ,  $p=0.05$ ). The apparent inverse correlation for GLU/ASP oxygens is not statistically significant ( $\text{CC}=-0.15$ ,  $p=0.09$ ).



**Figure S5** *B*-factor (**above**) and *B*<sub>Damage</sub> (**below**) of ASP and ASN O<sub>δ</sub> atoms of the six Nanao *et al.* (2005) proteins in two different dose states. The deterioration of the ASP side chain termini of ribonuclease A cannot be observed from the *B*-factor alone, but is clearly indicated by *B*<sub>Damage</sub>. The ASN oxygens remain unaffected.

## References

- Sheldrick, G. M. (2010). Acta Cryst. D66, 479–485.  
 Thorn, A. & Sheldrick, G. M. (2011). J. Appl. Cryst. 44, 1285–1287.