



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Human Cytomegalovirus variant peptides adapt by decreasing their total coordination upon binding to a T cell receptor

Georgios S.E. Antipas<sup>a,\*</sup>, Anastasios E. Germenis<sup>b</sup><sup>a</sup> Division of Materials Technology, National Technical University of Athens, Zografou Campus, Athens 15780, Greece<sup>b</sup> Department of Immunology & Histocompatibility, School of Medicine, University of Thessaly, Biopolis, Larissa 41110, Greece

## ARTICLE INFO

## Article history:

Received 8 May 2015

Received in revised form

3 July 2015

Accepted 13 July 2015

Available online 26 July 2015

## Keywords:

pMHC–TCR interactions

Class I MHC

CD8<sup>+</sup> Cytotoxic lymphocytes

Protein–protein interactions

## ABSTRACT

The tertiary structure of the native Cytomegalovirus peptide (NLV) presented by HLA-A2 and bound to the RA14 T cell receptor was used as a reference for the calculation of atomic coordination differences of both the NLV as well as of a number of singly substituted NLV variants in the absence of TCR. Among the pMHC complexes, the native peptide was found to exhibit the highest total coordination difference in respect to the reference structure, suggesting that it experienced the widest structural adaptation upon recognition by the TCR. In addition, the peptide on the isolated NLV–MHC complex was over-coordinated as compared to the rest of the variants. Moreover, the trend was found to account for a set of measured dissociation constants and critical concentrations for target-cell lysis for all variants in complexation with RA14: functionally, all variant peptides were established to be either weak agonists or null peptides, while, at the same time, our current study established that they were also under-coordinated in respect to NLV. It could, thus, be argued that the most ‘efficient’ structural adaptation upon pMHC recognition by the TCR requires of the peptide to undergo the widest under-coordination possible. The main structural characteristic which differentiated the NLV in respect to the variants was a the presence of 16 oxygen atoms (waters) in the former’s second coordination shell which accounted for over-coordination of roughly 100% and 30% in the O–O and C–O partials respectively. In fact, in the absence of second shell oxygens, the NLV

\* Corresponding author. Tel.: +30 210 7722037.

peptide was decidedly under-coordinated in respect to all of the variants, as also suggested by the C–C partial.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications table

|                            |  |
|----------------------------|--|
| Subject area               | Immunology, biochemistry, materials science, quantum chemistry                           |
| More specific subject area | Class I MHC, CD8 <sup>+</sup> Cytotoxic lymphocytes, protein–protein interactions        |
| Type of data               | Excel spreadsheet  |
| How data was acquired      | Data from crystallized tertiary structures was acquired from the Protein Data Bank (PDB) |
| Data format                | Text   |
| Experimental factors       | None   |
| Experimental features      | None   |
| Data source location       | Not applicable   |
| Data accessibility         | Data is with this article  |

## Value of the data

- On the basis of our previous findings which suggest that peptide-Major Histocompatibility Complex (pMHC) binding affinity with a T Cell Receptor (TCR) may be inferred from peptide tertiary structure [1,2], our current results highlight the possible inference of peptide immunological identity from its tertiary structure on the MHC, in the absence of complexation to a TCR. Peptides on the isolated pMHC complexes were substantially over-coordinated in respect to the (native) NLV-A2-RA14 complex. Over-coordination was owing to the oxygen partials.
- As shown in Fig. 3a, the native peptide appears to be able to undergo the widest possible under-coordination when recognized by a TCR.

## 1. Introduction

Several lines of evidence indicate that Cytomegalovirus (CMV), a ubiquitous  $\beta$ -herpesvirus that infects 60–90% of the population, is a driving force in age-related T cell immunosenescence. Briefly, in CMV-seropositive older adults, aging has been associated with large expansion of CMV-specific CD8<sup>+</sup> T cell clones and shrinkage of the T cell repertoire available for other antigens [3–7]. In individuals sharing the widespread HLA-A\*0201 allele (referred to as A2), CMV-specific CD8<sup>+</sup> T cells recognize the same epitope pp65<sub>495–503</sub> (NLVPMVATV), hereafter referred to as NLV [8,9].

Despite the great effort made to date, the founding mechanism of NLV immunodominance has not been resolved. However, studies thus far have indicated that antigen-driven selection is most likely the main parameter contributing to the predominant usage of public TCR by NLV-A2-specific CD8 T cells [10–12]. Motivated by our previously reported finding that pMHC-TCR binding avidity of the transcriptional regulatory protein of the human T-cell leukemia virus type 1 (HTLV-1), Tax, along with that of a number of its variants with A6 TCR was highly correlated to atomic coordination of peptide tertiary structure [1,2], here we investigate the potential correlation between atomic coordination of a number of NLV variant-MHC complexes in respect to that of the NLV-MHC and we additionally inquire

whether this correlation may also be indicative of the end effect of pMHC-TCR binding avidity of the same complexes as recognized by the RA14 TCR.

## 2. Materials and methods

### 2.1. Peptides

The current study considered a number of Cytomegalovirus (CMV) peptide (NLV) variants, all of which are presented by HLA-A2. Each variant was synthesized via a single residue substitution on the NLV [11]. In our work, the tertiary structures of the variants were compared to that of the native peptide (3GSN\_P), the tertiary structure of which was isolated from its complex with HLA-A2 bound to the RA14 T TCR (PDB accession code 3GSN). The pMHC complexes studied are listed in Table 1.

### 2.2. Calculation of pair correlation functions

For all peptides in Table 1, as well as for the native peptide, the Pair Distribution Function (PDF, symbolized as  $g(r)$ ) was calculated with a bin size equal to 0.1 Å, by

$$g(r) = \frac{1}{2\pi N r^2 \rho_0} \sum_{j=1}^N \sum_{i>j}^N \delta(r - r_{ij}) \quad (1)$$

where  $N$  is the number of peptide atoms,  $\delta$  is the Dirac delta function and  $\rho_0$  is the number density  $N/V$ ,  $V$  is the volume of the simulation box containing the peptide. In (1), if the species of the  $i$ th and/or  $j$ th are restricted, the  $g(r)$  calculated represents a partial (e.g. if all  $i$  atoms are restricted to C the PDFs computed would be the carbon partials; additionally if the  $j$  atoms are also restricted to, e.g., oxygen, the partial computed would be the C–O), otherwise it is the total PDF. The Radial Distribution Function (RDF, symbolized as  $R(r)$ ), was then calculated as

$$R(r) = 4\pi r^2 \rho_0 g(r) \quad (2)$$

and integrated to estimate the cumulative atomic coordination,  $n^{r_1}$ , of any atom within a sphere of radius  $r_1$  as follows

$$n^{r_1} = \int_0^{r_1} R(r) dr = 4\pi \rho_0 \int_0^{r_1} g(r) r^2 dr \quad (3)$$

Hence the running difference between the coordination of the peptide on the NLV-A2-RA14 complex and each of the peptides in Table 1 was calculated. These running differences in cumulative coordination are presented in Fig. 1. It follows from the definition of the PDF (Eq. 1) that both total and partial RDF and coordination may be calculated. All calculations of PDF, RDF and coordination were performed with the PRDF program [13–16].

### 2.3. Side chain mobility and Density Functional Theory calculations

In the case of the 3GSO\_P, 3GSV\_P, 3GSU\_P and 3GSX\_P complexes increased side-chain mobility resulted in multiple atomic conformations as recorded via X ray diffraction in the work by Gras et al. [11]. Here, for the calculation of the pair correlation functions we created unique structural models (see ‘Number of different pMHC models’ in Table 1) for each peptide with multiple atom positions via the following two separate routes:

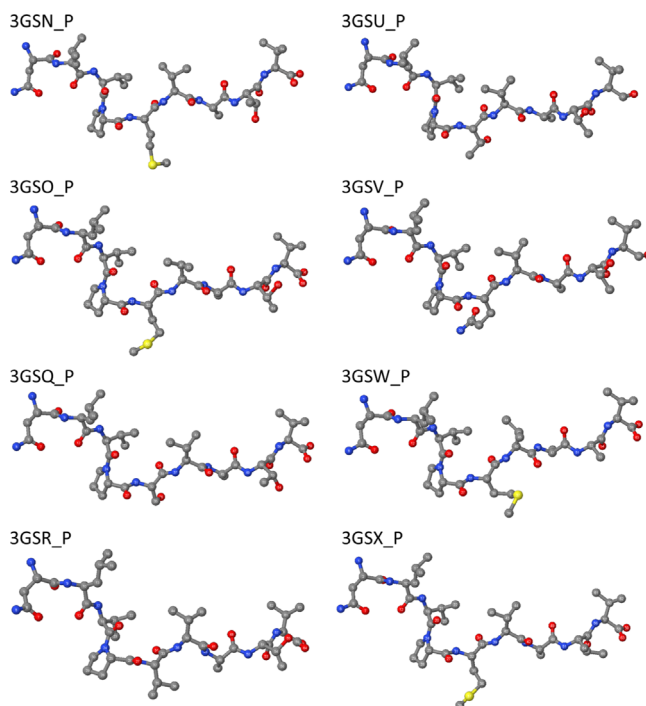
- a) If an atom's multiple positions differed by less than its covalent radius, all positions but one were discarded as small PDF histogram differences do not contribute appreciably to the total coordination of the peptide. In cases where an atom's possible positions were separated by more than the covalent radius, we produced different peptide models having both included and excluded second coordination shell oxygen atoms and calculated coordination for each model. All models studies are listed in Table 2.

**Table 1**

All peptides were loaded on HLA-A\*0201.

| PDB ID | Peptide designation | Peptide sequence | Number of 2nd shell O | Number of different pMHC models | $K_D$ | $EC_{50}$            |
|--------|---------------------|------------------|-----------------------|---------------------------------|-------|----------------------|
| 3GSO   | 3GSO_P              | NLVPMVATV        | 16                    | 96                              | 27.7  | $5.10^{-2} \pm 0.01$ |
| 3GSV   | 3GSV_P              | NLVPOVATV        | 6                     | 3                               | 57.5  | $10^2 \pm 20$        |
| 3GSQ   | 3GSQ_P              | NLVPSVATV        | 7                     | –                               | n.b.  | $5.10^2 \pm 20$      |
| 3GSR   | 3GSR_P              | NLVPSVATV        | 9                     | 2                               | 60.2  | $5.10^2 \pm 50$      |
| 3GSU   | 3GSU_P              | NLVPTVATV        | 9                     | 3                               | 58.2  | $10^3 \pm 30$        |
| 3GSW   | 3GSW_P              | NLVPMVAIV        | 7                     | –                               | n.b.  | $> 10^4$             |
| 3GSX   | 3GSX_P              | NLVPMVAIV        | 6                     | 3                               | n.b.  | $> 10^4$             |

All PDB entries listed refer to pMHC complexes. For each of the variants the amino acid substituted on the original NLV sequence (i.e. 3GSO\_P) is underlined. Each peptide was surrounded by second coordination shell oxygen atoms (they are hence designated as '2nd shell O' in the table). The number of different models arising for each peptide, due to substantial side-chain mobility during crystallization [11], is listed in the fifth table column. For 3GSO\_P multiple positions involved the following atoms (notation is the atom species followed by its index in the PDB entry, multiple positions for the same species are separated by '/'): C36/C37, S38/S39, C40/C41 on residue 5, C47/C48/C49/C50 on residue 6, C63/C64 on residue 8 and O61/O62 on residue 8. 3GSV\_P and 3GSU\_P involved C57/C58 and O59/O60 on residue 8. Gras et al. [11] also measured the functional avidity of each of the variant pMHC complexes upon recognition by RA14 TCR (these pMHC-TCR structures were not crystallographed) and these values are presented as dissociation constants ( $K_D$ ) and peptide concentration necessary to achieve half maximal target-cell lysis ( $EC_{50}$ ) (notation 'n.b.' signifies that no pMHC-TCR binding was detected).



**Fig. 1.** The stereochemistries of an indicative subset of peptides considered in the current study. In all peptides shown, second coordination shell oxygen atoms have been excluded. The peptides are (also see model designation in Table 2): A1 (3GSO\_P), A4 (3GSQ\_P), A5 (3GSR\_P), A7 (3GSU\_P), A10 (3GSV\_P), A13 (3GSW\_P) and A14 (3GSX\_P). Reference peptide 3GSN\_P is also shown.

- b) To further account for the structural complexity due to side-chain mobility in the 3GSO\_P peptide, all of its atoms were first saturated with hydrogens assuming a neutral pH (peptide N and C termini involved  $NH_3^+$  and  $COO^-$  groups, respectively). The side-chain atoms in question along with the entire H species were then allowed to relax as charge-neutral zwitterions via closed-shell

**Table 2**  
Peptide models used in the coordination calculations.

| Peptide designation | Model                | 2nd Shell O |          |
|---------------------|----------------------|-------------|----------|
|                     |                      | Excluded    | Included |
| 3GSO_P              | 36_38_40_49_48_63_61 | A1          | B1       |
| "                   | 37_39_41_50_48_64_61 | A2          | B2       |
| "                   | BLYP_TZ2P            | A3          | B3       |
| 3GSQ_P              | –                    | A4          | B4       |
| 3GSR_P              | 56                   | A5          | B5       |
| "                   | 57                   | A6          | B6       |
| 3GSU_P              | 57_55                | A7          | B7       |
| "                   | 58_55                | A8          | B8       |
| "                   | 58_56                | A9          | B9       |
| 3GSV_P              | 59_57                | A10         | B10      |
| "                   | 60_57                | A11         | B11      |
| "                   | 60_58                | A12         | B12      |
| 3GSW_P              | –                    | A13         | B13      |
| 3GSX_P              | 58_56                | A14         | B14      |
| "                   | 58_57                | A15         | B15      |
| "                   | 59_57                | A16         | B16      |

All models refer to peptide tertiary structure isolated from the pMHC complexes. Peptide designation follows from Table 1 and Model refers to the details given in the caption of Table 1. Model designation starts with 'A' and 'B' if second coordination shell oxygen atoms were included in and excluded from the structure, respectively.

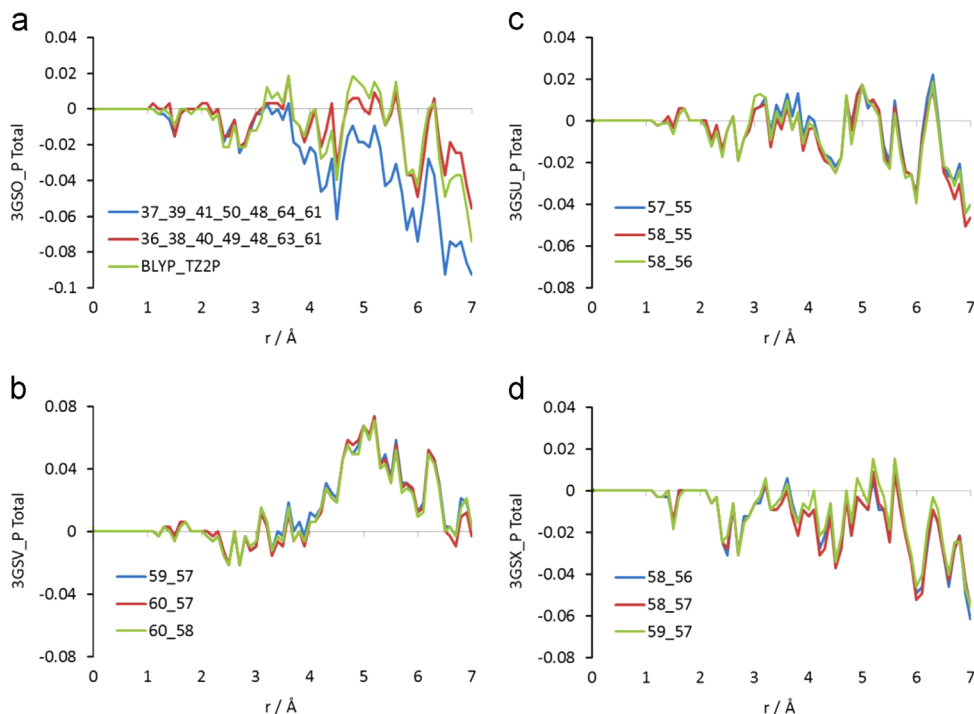
Density Functional Theory (DFT), while keeping the rest of the peptide atoms immobile in their crystallographed positions. The DFT relaxed model for 3GSO\_P is designated as 'A3' in Table 2.

DFT calculations were performed with the Amsterdam density functional (ADF) program [17,18]. Electron exchange and correlation were addressed within the generalized gradient approximation (GGA) by the BLYP [19,20] functional. Single-electron wavefunctions were expanded using the TZ2P uncontracted Slater-type orbital (STO) basis set, (a triple- $\zeta$  basis set with two sets of polarization functions) for all atoms. Calculations were all-electron for the H species, while for C, N and O core electrons were frozen inclusive of the 1s shell whereas for S the frozen core included 2p. During relaxation, aufbau occupations were always imposed. Self-consistent field (SCF) convergence invariably required use of the Augmented Roothaan-Hall Direct Inversion Iterative Subspace (ADIIS) scheme [21].

### 3. Brief discussion

Indicative peptide stereochemistries are shown in Fig. 1. In the absence of second shell O atoms, the native peptide, 3GSO\_P, is under-coordinated in respect to the reference (also see Fig. 3b) and may not be differentiated from the other peptides exhibiting increased side-chain mobility (3GSU\_P, 3GSV\_P and 3GSX\_P in Fig. 1). Additionally, as shown in Fig. 2, the variance of the total coordination of the 3GSO\_P models varies is substantially higher, roughly four times that of the other three peptides associated with mobile side-chains. Interestingly, the coordination difference of the DFT-relaxed 3GSO\_P model from the reference falls well within the range of 3GSO\_P models. It might, thus, be envisaged that the 'flexibility' shown by 3GSO\_P to adapt to recognition by the TCR rests on the substantially higher number of second shell O atoms – almost twice as many O atoms in comparison to the rest of the variants, also see Table 1. Due to their abundance, second shell O atoms, which represent water molecules coordinated around the peptide might provide adequate 'cushioning' to the NLV such that the peptide is able to adjust to a wide variety of TCR's via expulsion of waters during recognition [22].

Certainly, inclusion of the second shell O atoms in peptide structures produced a trend of total coordination vs. functional avidity of the peptide in complexation with the TCR which was

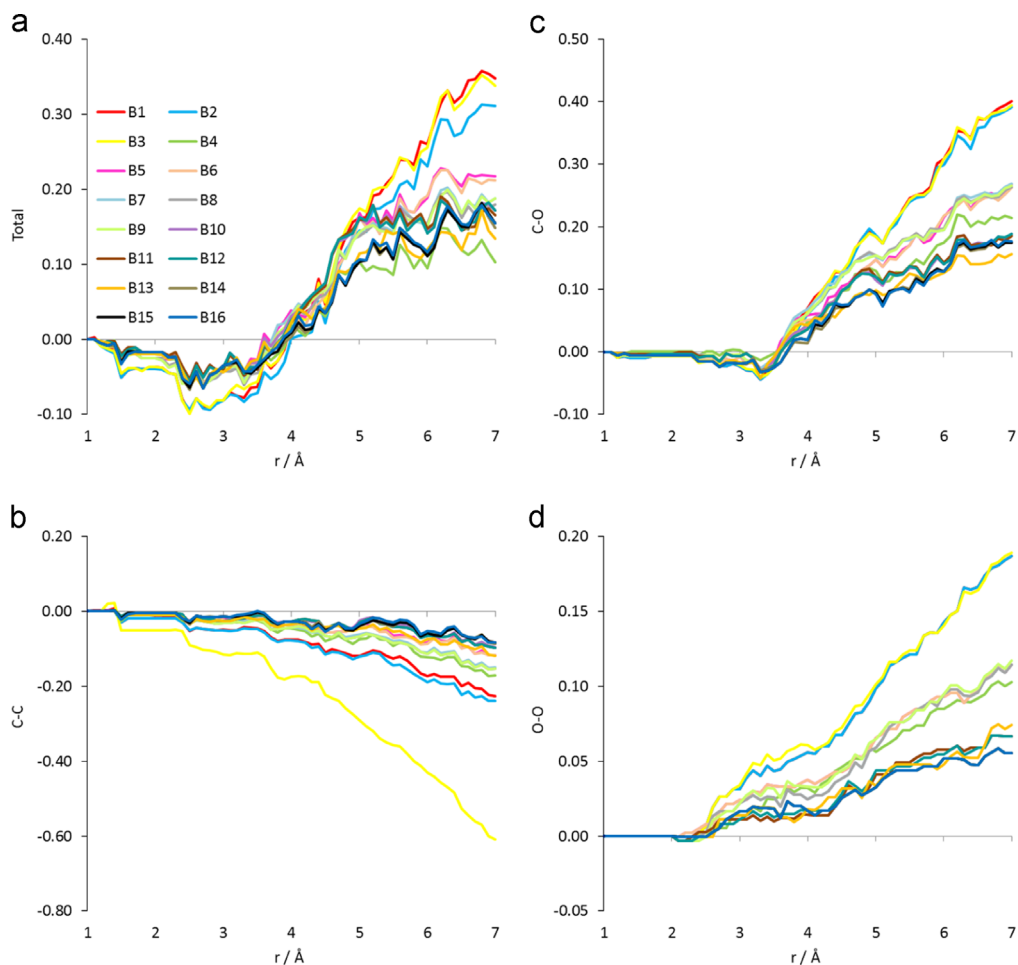


**Fig. 2.** Evolution of the difference between the cumulative total coordination for each of the variants which exhibited increased side-chain mobility, in respect to the native peptide. Coordination has been calculated for unprotonated structures excluding second shell O atoms. (a) Of the conformations possible for 3GSO\_P, the two models which yielded the highest and lowest coordination in respect to the native peptide are shown. Additionally, coordination of the model for which side chain atoms were allowed to undergo DFT relaxation is included. (b), (c) and (d) Coordination differences of the three possible models for each of the 3GSV\_P, 3GSU\_P and 3GSX\_P peptides respectively. In the graphs, the data series designation involves the indexes (as they appear in the original PDB entry) of the mobile side-chain atoms present in each model.

reminiscent of our results on atomic coordination of pMHC-TCR complexes in the case of Tax and a number of Tax-derived variants [1,2], i.e. the agonist peptide is over-coordinated in respect to the less agonist/antagonist. As depicted in Fig. 3a, all three 3GSO\_P models were over-coordinated in respect to the reference as well as compared to the rest of the variants. In fact, 3GSO\_P over-coordination in respect to the variants was almost exclusively due to the C–O and, principally, the O–O partials by roughly 60% (see Fig. 3c and d respectively) as compared to the highest-coordinated of the variants.

#### 4. Note on the data files

All peptide structures considered in the current study have been included in .xyz format in the Structures.rar file. The pair correlation data underlying to this work are included in PRDF.rar in comma delimited format. Data comparison is provided in PRDF.xls, which comprises total and partial PDF, RDF and coordination ( $RDF(r)dr$ ) data for each of the peptides studied, in respect to interatomic distance,  $r$  (Å). In each of the tabs, the first line of every column represents the peptide number density (atoms/Å<sup>3</sup>) while the second line is the peptide designation, corresponding to Table 1. The sequence of peptides is the same in each tab. In the set of tabs named 'PDF XXX', the peptide data have been calculated via Eq. (1), while data in the tabs named 'RDF XXX', the peptide data have been calculated via Eq. (2). The  $RDF(r)dr$  XXX have been calculated via Eq. (3). and to estimate the cumulative coordination for a peptide structure, the running sum of  $RDF(r)dr$  may be calculated as shown in the



**Fig. 3.** Total and partial cumulative coordination differences between the unprotonated tertiary structures of each variant (Table 1) and the tertiary structure of the peptide on NLV-A2-RA14 complex (PDB ID 3GSN). All variants include second shell oxygen atoms. (a) Total coordination and coordination partials (b) C-C, (c) C-O and (d) O-O.

example tab 'RDF( $r$ )dr Total' and coordination differences in respect to the column named 'Reference' may then be calculated also as exemplified.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2015.07.019>.

## References

- [1] G.S.E. Antipas, A.E. Germeis, The quantum chemical causality of pMHC-TCR biological avidity: peptide atomic coordination data and the electronic state of agonist N termini, *Data Brief* 3 (2015) 180–184.
- [2] G.S.E. Antipas, A.E. Germeis, Quantum chemical calculations predict biological function: The case of T cell receptor interaction with a peptide/MHC class I, *Front. Chem.* 3 (2015) 3.

- [3] N. Khan, et al., Cytomegalovirus seropositivity drives the CD8 T cell repertoire toward greater clonality in healthy elderly individuals, *J. Immunol.* 169 (4) (2002) 1984–1992.
- [4] S.R. Hadrup, et al., Longitudinal studies of clonally expanded CD8 T cells reveal a repertoire shrinkage predicting mortality and an increased number of dysfunctional cytomegalovirus-specific T cells in the very elderly, *J. Immunol.* 176 (4) (2006) 2645–2653.
- [5] R. Vescovini, et al., Massive load of functional effector CD4<sup>+</sup> and CD8<sup>+</sup> T cells against cytomegalovirus in very old subjects, *J. Immunol.* 179 (6) (2007) 4283–4291.
- [6] B. Pourgheysari, et al., The cytomegalovirus-specific CD4<sup>+</sup> T-cell response expands with age and markedly alters the CD4<sup>+</sup> T-cell repertoire, *J. Virol.* 81 (14) (2007) 7759–7765.
- [7] R. Vescovini, et al., Intense antiextracellular adaptive immune response to human cytomegalovirus in very old subjects with impaired health and cognitive and functional status, *J. Immunol.* 184 (6) (2010) 3242–3249.
- [8] D.J. Diamond, et al., Development of a candidate HLA A\* 0201 restricted peptide-based vaccine against human cytomegalovirus infection, *Blood* 90 (5) (1997) 1751–1767.
- [9] S. Mackinnon, et al., Characterization of human cytomegalovirus peptide-specific CD8 T-cell repertoire diversity following in vitro restimulation by antigen-pulsed dendritic cells, *Blood* 99 (1) (2002) 213–223.
- [10] V. Venturi, et al., The molecular basis for public T-cell responses? *Nat. Rev. Immunol.* 8 (3) (2008) 231–238.
- [11] S. Gras, et al., Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope, *J. Immunol.* 183 (1) (2009). 430–7.
- [12] J.B. Reiser, et al., Crystallization and preliminary X-ray crystallographic characterization of a public CMV-specific TCR in complex with its cognate antigen, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 65 (Pt 11) (2009) 1157–1161.
- [13] G.S.E., Antipas, PRDF: software for the calculation of atomic pair correlation and short-range ordering, (<http://users.ntua.gr/gantipas/prdf/>), NTUA, Athens, 2014.
- [14] G.S.E. Antipas, A concise methodology for the estimation of elemental concentration effects on mesoscale cohesion of non-ferrous covalent glasses: the case of  $\text{Se}_{(80-x)}\text{Ge}_{(20-x)}\text{In}_{x=0.5,10,15}$ , *Data Brief* 4 (2015) 257–265.
- [15] G.S.E. Antipas, Short range order and stability of a mechanically alloyed  $\text{Cr}_{25}\text{Nb}_{75}$  alloy determined by total scattering and first principles, *Mater. Res.* 17 (6) (2014) 1677–1685.
- [16] G.S.E. Antipas, Molecular orbital interactions in glass-forming  $\text{Zr}_{70}\text{Cu}_{30}$  liquid quasicrystals, *J. Alloy. Compd.* 578 (2013) 565–570.
- [17] SCM, Amsterdam Density Functional (ADF) program, Theoretical Chemistry, (<http://www.scm.com/ADF/>), Vrije Universiteit, Amsterdam, 2014.
- [18] G. te Velde, et al., Chemistry with ADF, *J. Comput. Chem.* 22 (9) (2001) 931–967.
- [19] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A* 38 (6) (1988) 3098–3100.
- [20] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. A* 37 (2) (1988) 785–789.
- [21] X. Hu, W. Yang, Accelerating self-consistent field convergence with the augmented Roothaan–Hall energy function, *J. Chem. Phys.* 132 (5) (2010) 054109.
- [22] W.W. Schamel, M. Reth, The TCR binding site does move, *Proc. Natl. Acad. Sci. USA* 104 (42) (2007). 16398–9.