GCB2011 Junior Paper

Modeling the Variable Regions of T-Cell Receptors

Thomas Hoffmann* and Iris Antes

Department of Life Sciences and Center for Integrated Protein Science Munich, Technical University Munich, 85354 Freising-Weihenstephan, Germany
E-mail: hoffmant@wzw.tum.de

Heterodimeric T-cell receptors (TCR) bind and recognize pathogenic molecules presented by major histocompatibility complex (MHC) molecules. The relative orientation of the two TCR chains can differ depending on their type. Here we present a comprehensive analysis of this phenomenon based on the current set of known TCR X-ray structures. During the analysis the structures could be grouped into several clusters. Based on these structural results, we developed a method to predict the TCR chain association. Such a prediction is e.g. important to reduce complications in adoptive T-cell therapy. Furthermore, the orientation might be important for the TCR:peptide:MHC (TCR-p-MHC) complex formation.

1 Introduction

T-cells play a major role in the adaptive immune response. Heterodimeric T-cell receptor (TCR) molecules distinguish between self-peptides and pathogenic nonself-peptides. Peptides are presented on the surface of cells by major histocompatibility complex molecules (MHC). The capability of the immune system to recognize many different peptide-MHC complexes (pMHC) is achieved by a vast variety of different TCRs. The T-cell repertoire was estimated to 10^{12} for one human individual[1]. Next to the vdj-recombination of TCR genes, the pairing of different α- and β- chains contributes to the TCR variety.
This work focuses on the question, if the TCR chain pairing occurs in different orientations. Different orientations of the TCR chains might be crucial for the reliable prediction of chain pairing and the prediction of the binding behavior of a TCR to a pMHC complex. For example, the misspairing of TCR chains in transduced T-cells used in cancer therapy can lead to autoreactive lymphocytes with lethal consequences[2].
The determination of TCR inter-chain geometries is complicated by the fact that structural data is only available for a small subset of the vast variety of TCRs and that the TCRs for which structural data is available, differ considerably in their loop structure and chain length, rendering the location of common conserved structural elements difficult. To deal with these complications, a method was developed, which transfers the TCR chains into a unified geometric framework. Based on the transformed structures we performed an analysis of the TCR structure geometries and developed a procedure for modeling new TCR chain combinations.

2 Material and Methods

2.1 Data Set

A set of 106 X-ray crystal structures was acquired from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/)[3]. The used structures contain bound and unbound TCRs from *H. sapiens* and *Mus musculus*. Each crystallographically independent molecule in the asymmetric unit (IMAU) was treated as a separate structure, leading to a total amount of 163 different TCR complexes. This set is further referred to as *S*.

2.2 Superpositioning and Cuboid Placement

First, all TCR structures were reduced to their variable binding domains (V). Then we defined unified cuboids for each V-domain of the different TCR chains. The cuboid templates (CT^{α} and CT^{β}; Figure 1) comprise a cuboid of the size of the extent of the 2bnu[5] α- or β-chain V domain and a reduced set of the 2bnu α- or β-chain V domain framework residues. The reduced set was defined to allow for a robust superpositioning of the experimental structures during our modeling procedure. For the superpositioning the tool DALI[4] was used together with the defined subset of V-framework residues as templates (Figure 1). This tool does not take sequence conservation into account and is purely structure based. To define the subset, in a preparatory step the structures of each chain were superimposed separately. For this purpose all loops and turns were removed and the template residues were determined iteratively from the remaining residues, such that the set of mapped residues used as superpositioning anchors in DALI was converged and the variance of the backbone root mean square deviation (RMSD) over all superposed structures was low. After the subsets were identified, the following procedure was used to superpose the combined chains and to place the cuboids. First, the αV:βV-complexes were superimposed based on their α-chains using the above defined α-subset and the tool DALI. All structures were superimposed to the high resolution (1.4 Å) structure with the PDB ID 2bnu. This step leads to a set of TCR-structures, which is further referred to as S^{α} and a corresponding set of cuboid templates (CT^{α}), containing cuboids which were placed around the α-chains based on the positions of the α-subset residues. Second, the same procedure was used to align cuboid templates (CT; Figure 1) additionally around each β-chain contained in the set S^{α} according to the relative position of the β-chains towards their paired, superposed α-chains resulting in a set of cuboid templates around the β-chains (CT^{β}). This step results in the set C consisting of the β-chain cuboids (CT^{β}) and the corresponding β-V-domain structures.

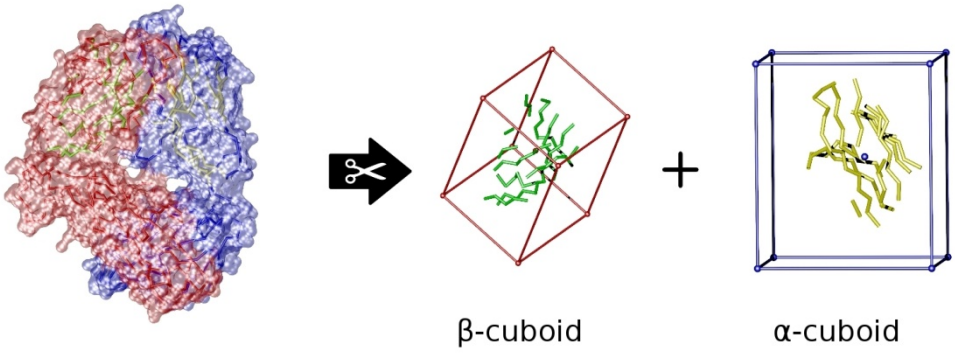


Figure 1: *Preparation of the reference structure (PDB ID 2bnu[5]) for superpositioning. The α-cuboid (yellow/blue) and β-cuboid (green/red) shown demonstrate the cuboid and the region used for superpositioning after loops and flexible residues were removed.*

2.3 Comparing the relative β-Chain Geometries

We made use of two different approaches to determine the similarity *d(i,j)* between two chain geometries *i,j* of the set *C*, consisting of the relative β-chain geometries of the TCR structures superpositioned with respect to the α-chains (set S^{α}). For the first approach, we simply computed the RMSD of the eight cuboid vertices and the cuboid center:

$$d_R(i,j) = RMSD(i,j) = \sqrt{\frac{1}{9}\sum_{k=1}^{9}((x_{ki}-x_{kj})^2+(y_{ki}-y_{kj})^2+(z_{ki}-z_{kj})^2)} \qquad (1)$$

For the second approach, we computed the Euler angles for each cuboid geometry with respect to a reference coordinate system. The calculation was implemented using the GNU generic math template library; all angles were computed in xyz-order. The reference coordinate system was chosen to be the coordinate system of the 2bnu structure. The similarity between two geometries we defined as the

Euclidean distance of the Euler angles:

$$d_E(i,j) = \sqrt{(\Phi_i - \Phi_j)^2 + (\Psi_i - \Psi_j)^2 + (\Theta_i - \Theta_j)^2} \qquad (2)$$

The distance matrices D_R and D_E were clustered hierarchically by average[6]. For each cluster the geometry with the lowest average distance to all other members in the cluster was chosen and further referred to as cluster representative (CR).

2.4 Modeling of Chain Pairings

The method (Figure 2) is based on the resulting structures and cuboids from Section 2.2 and the clusters identified in Section 2.3 and consists of four steps. First, for a given target TCR sequence the most homologous α- and β-chains are identified from the set *S* of available TCR structures. Second, the chains of the identified template structures are assembled according to the geometries of all CR structures, leading to a set of several templates with differing geometries. Third, the tool MODELLER[7] is used to model the backbones of the target sequences based on the CR template structures. Fourth, all side chains are placed with the IRECS-algorithm[8] available in the DynaCell software package. For each CR geometry the according homology model is evaluated using the soft-core scoring function ROTA[8, 9]:

$$\Delta E_{bind} = ROTA(\alpha\beta) - (ROTA(\alpha) + ROTA(\beta)) \qquad (3)$$

For evaluation of the method, we remodeled a subset of *S* containing only MHC-bound TCRs (83 different structures). In this remodeling procedure, we omitted step 3 (MODELLER), since we were interested in the performance of the αβ-chain placement procedure, rather than the accuracy of the MODELLER tool.

3 Results and Discussion

3.1 Analysis of the Experimental Structures

Analysis of the superpositioned structures in set S^{α} showed that the relative positions of the α- and β-chains of the variable domains of the TCRs differ considerably with respect to each other (Figure 3a and b). In Figure 3a it can be observed that if the central β-sheets of the α-domain are superposed very well, the backbone positions of the corresponding β-chains differ significantly (d_E values up to 30°). Therefore, the two TCR binding domains can adopt different orientations (Figure 3). To analyze these differences we clustered the structures according to their positional and angular deviations in the β-chains. This clustering led to eight to ten clusters using clustering limits of about 2.5 Å or 8°, respectively. Both clustering criteria led to similar clusters. This shows that the clustering is robust. In addition TCRs of the same type but from different X-ray structures were placed in the same cluster, indicating that the observed phenomenon is not caused by the variation of the crystallographic conditions and that the different orientations have to be considered for the creation of homology models of TCR- and TCRpMHC-complexes. Another interesting point is that structures from human and mouse are found in the same clusters, no differences were observed in their clustering behavior.

3.2 Remodeling of α-β Chain Association

Our long-term goal is to develop a homology modeling method with which the different chain geometries of TCRs can be reliably predicted. One of the main obstacles is the gap between the number of known TCR structures and the vast variety of possible receptors. However, the problem can be reduced by combining structures of individual TCR domains from different experimental structure files to obtain templates for homology modeling. Our method allows assembling arbitrary αV- and βV-domains into a single modeled structure based on the different CR geometries obtained during the analysis step and is thus well suited for this purpose.

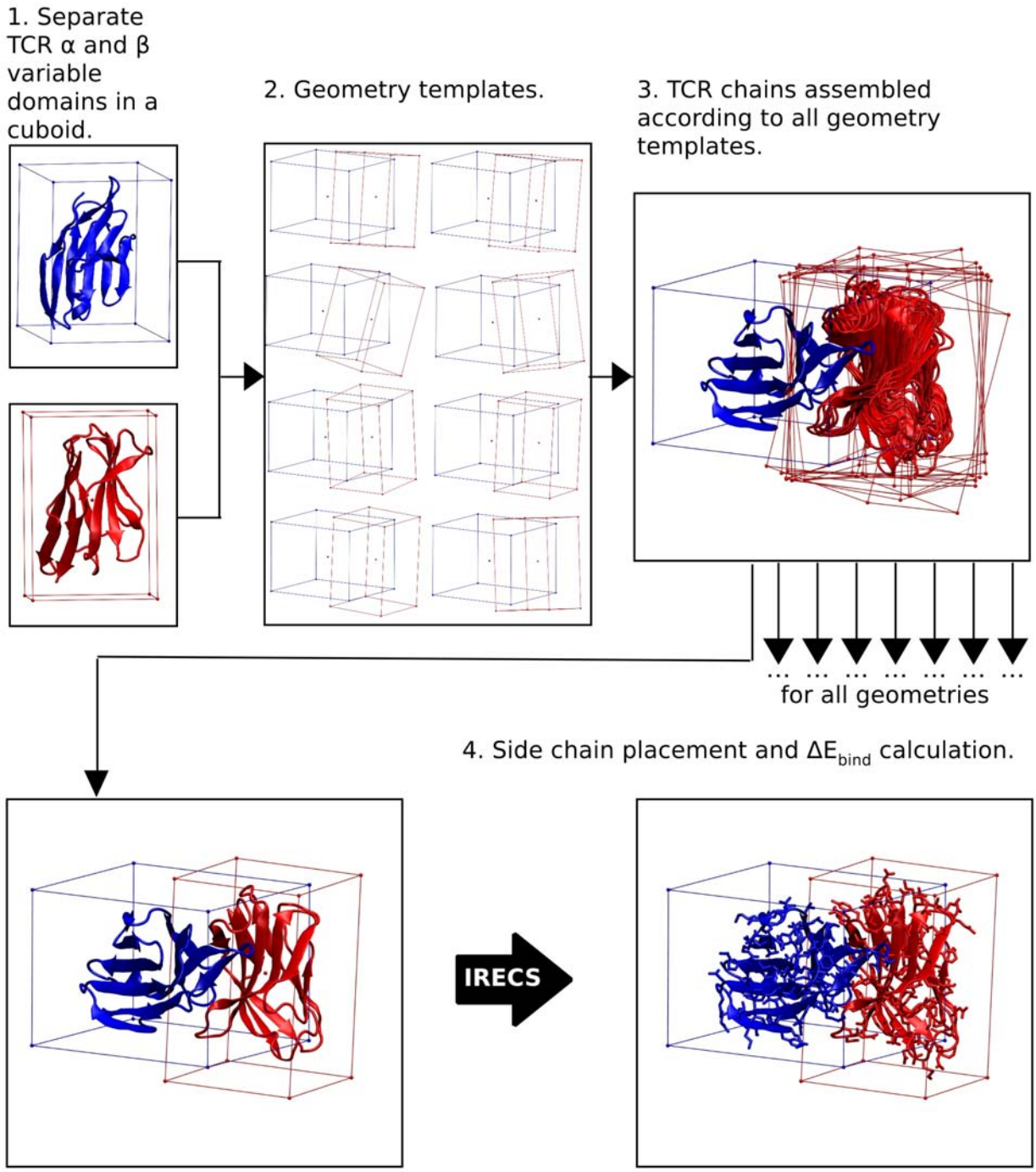


Figure 2: *Remodeling procedure. The separated α- and β- V domains are surrounded each by cuboids (1). According to eight different geometry templates (2) the two chains are assembled according to the cuboid geometries. This step results in eight different orientations of the two chains (3). For each of the eight assemblies, the side chains are replaced using the IRECS-algorithm*[8] *and the binding energy is determined.*

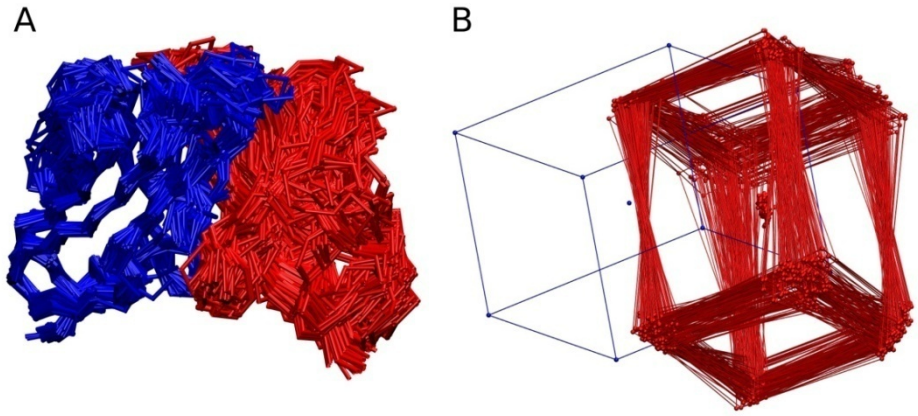


Figure 3: *Variable regions of the examined set of TCR structures superposed onto the α-chains (blue). The corresponding β-chains are shown in red. (A) shows the actual α-superposed backbones and (B) the corresponding cuboids.*

We evaluated the method on a subset of 83 MHC-bound TCR structures. 73% of the structures were assigned to the correct geometry CR, using the two best ranking ΔE_{bind} values obtained with the ROTA scoring function, which corresponds to 19 of 22 different receptor types in the evaluation set. Only three receptor types showed a bad correlation between their ΔE_{bind} values and the distances to their native CR. Structural analysis of these models revealed rough surfaces of the inter-chain binding interfaces, leading to numerous side chain clashes which were too large to be refined properly by the IRECS algorithm. We found a strong correlation between the number of clashes and high values of the predicted association energies. Therefore, although our method already shows an overall good performance, we are currently working on the improvement of the accuracy of our prediction using the OPMD[10] approach for further refinement of the modeled structures. The method then could be used to predict the pairing and misspairing of transduced TCRs or could be extended to TCR-p-MHC complexes for T-cell epitope prediction.

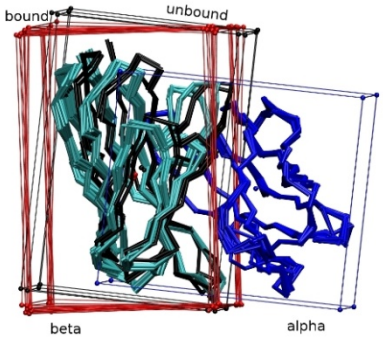
3.3 Differences between bound and unbound TCRs

Based on the analysis and clustering of the experimental structures we compared the structural features of the unbound and bound TCR structures of the same type to analyze the influence of MHC binding on the overall TCR structures.

Comparing bound and unbound TCRs of the same type, we found that in most of the cases the orientations of the unbound TCRs slightly differ from the bound TCRs. Nevertheless, bound and unbound structures of the same TCR type tend to cluster in the same or in closely related clusters. Comparing all examined structures of bound and unbound TCRs, in general the differences in the β-chain orientations are considerably larger for the unbound TCRs.

In Figure 4a the differences between the bound and the unbound 1G4 TCR structures is illustrated. The two unbound 1G4 orientations were derived from two different crystal structures, whereas the eight bound orientations were obtained from seven different crystals. Notably, the two unbound orientations differ only by 2.5 degrees, but have an average difference of 8° from the cluster of bound representatives. On the other hand, all bound 1G4 TCR structures are very similar.

This indicates a shift in the relative orientation of the two chains upon binding of the TCR to the peptide-MHC complex; however, the differences between the bound structures of different TCRs are considerably larger than this shift.



	U1	U2	B1a	B1b	B2	B3	B4	B5	B6	B7
U1	0.00	2.51	7.17	9.73	7.82	8.67	9.00	9.29	6.66	7.90
U2	2.51	0.00	5.74	8.88	7.83	8.49	8.85	8.67	6.11	7.17
B1a	7.17	5.74	0.00	3.60	4.57	4.58	4.87	3.74	2.52	2.43
B1b	9.73	8.88	3.60	0.00	3.32	2.55	2.60	0.96	3.10	1.89
B2	7.82	7.83	4.57	3.32	0.00	0.99	1.24	2.49	2.17	2.40
B3	8.67	8.49	4.58	2.55	0.99	0.00	0.38	1.74	2.50	2.21
B4	9.00	8.85	4.87	2.60	1.24	0.38	0.00	1.78	2.85	2.47
B5	9.29	8.67	3.74	0.96	2.49	1.74	1.78	0.00	2.65	1.52
B6	6.66	6.11	2.52	3.10	2.17	2.50	2.85	2.65	0.00	1.33
B7	7.90	7.17	2.43	1.89	2.40	2.21	2.47	1.52	1.33	0.00

Figure 4. Comparing bound and unbound TCRs of the same type (1G4). Bound and unbound chains adopt similar orientations, respectively. The two unbound TCRs are taken from two different crystal structures. The eight bound TCRs are taken from seven different crystal structures. Left: The superimposed α-chains and the α-cuboid are shown in blue. Bound β-cuboids/chains are shown in red/cyan. Unbound α-cuboids/chains are shown in black. Right: Pairwise Euler-angle distances between the unbound (U) and bound (B*) β-chain structures.*

4 Conclusions

We developed a method to evaluate the relative orientation of β- and α-chains in T-cell receptor binding domains. Our studies, which are based on the currently available set of TCR structures in the PDB database, show significant differences in the relative orientation of the single chains of the TCR binding domain. Based on these results, we developed an approach to build and evaluate homology models of arbitrary TCR chain pairings. The approach was evaluated for the αβ-chain orientation remodeling procedure using a set of 83 bound TCR structures. In this evaluation the correct orientation was found for 73% of the structures. Analysis of the unbound versus MHC-bound TCR structures revealed that the differences in chain orientation within the unbound data set were larger than in the bound data set. A possible explanation would be that a certain orientation is induced and stabilized by the binding to a MHC molecule. It can be speculated, that the loss of mobility of the two binding domains plays a role in the signaling pathway of T-cells. The signal transduction of T-cells is not yet fully understood, since no common conformational changes within the single TCR chains upon binding of an antigen can be observed. Through our analysis and modeling procedures we hope to be able to further investigate and elucidate these open questions in more detail.

References

1. T. P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky, *A direct estimate of the human alphabeta T cell receptor diversity* , Science, 286, no. 5441, 958–961, Oct 1999.
2. G. M Bendle, C. Linnemann, A. I. Hooijkaas, L. Bies, M. A. de Witte, A. Jorritsma, A. D. M. Kaiser, N. Pouw, R. Debets, E. Kieback, W. Uckert, J.-Ying Song, J. B. A. G. Haanen, and T. N. M. Schumacher, *Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy*., Nat Med, 1, no. 5, 565–70, May 2010.
3. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *The Protein Data Bank.* , Nucleic Acids Res, 28 , no. 1, 235–242, Jan 2000.
4. H. Hasegawa and L. Holm, *Advances and pitfalls of protein structural alignment*., Curr Opin Struct Biol, 19, no. 3, 341–348, Jun 2009.
5. J.-L. Chen, G. Stewart-Jones, G. Bossi, N. M. Lissin, L. Wooldridge, E. M. L. Choi, G. Held, P. R. Dunbar, R. M. Esnouf, M. Sami, J. M. Boulter, P. Rizkallah, C. Renner, A. Sewell, P. Anton van der Merwe, B. K. Jakobsen, G. Griffiths, E. Y. Jones, and V. Cerundolo, *Structural and kinetic basis for heightened immunogenicity of T cell vaccines.*, J Exp Med, 201, no. 8, 1243–1255, Apr 2005.
6. R. R. Sokal and C. D. Michener, *A statistical method for evaluating systematic relationships*, University of Kansas Scientific Bulletin, 28, 1409–1438, 1958.

7. A. Sali, Comparative protein modeling by satisfaction of spatial restraints., Mol Med Today, 1, no. 6, 270–277, Sep 1995.
8. C. Hartmann, I. Antes, and T. Lengauer, *IRECS: a new algorithm for the selection of most probable ensembles of side-chain conformations in protein models*., Protein Sci, 16, no. 7, 1294–1307, Jul 2007.
9. C. Hartmann, I. Antes, and T. Lengauer, *Docking and scoring with alternative side-chain conformations.*, Proteins, 7, no. 3, 712–726, Feb 2009.
10. I. Antes, *DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility.*, Proteins, 78, no. 5, 1084–1104, Apr 2010.