ICM-DISCO Docking by Global Energy Optimization With Fully Flexible Side-Chains

Juan Fernández-Recio,[1] **Maxim Totrov,**[2] **and Ruben Abagyan**[1*]

[1]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California*
[2]*Molsoft, LLC, La Jolla, California*

ABSTRACT **The ICM-DISCO (Docking and Interface Side-Chain Optimization) protein–protein-docking method is a direct stochastic global energy optimization from multiple starting positions of the ligand. The first step is performed by docking of a rigid all-atom ligand molecule to a set of soft receptor potentials precalculated on a 0.5 Å grid from realistic solvent-corrected force-field energies. This step finds the correct solution as the lowest energy conformation in almost 100% of the cases in which interfaces do not change on binding. The second step is needed to deal with the induced changes and includes the global optimization of the interface side-chains of up to 400 best solutions. The CAPRI predictions were performed fully automatically with this method. Available experimental information was included as a filtering step to favor expected docking surfaces. In three of the seven proposed targets, the ICM-DISCO method found a good solution (>50% of correct contacts) within the five submitted models. The procedure is global and fully automated. We demonstrate that the algorithm handles the induced changes of surface side-chains but is less successful if the backbone undergoes large-scale rearrangements. Proteins 2003;52:113–117.** © 2003 Wiley-Liss, Inc.

INTRODUCTION

The biggest challenge in computational protein docking is to combine high-accuracy energy calculations, speed and sampling power, and the ability to handle induced conformational changes at the interface. Fortunately, these changes are often small, which suggest that the rigid body-docking procedure can be used to narrow down the list of possible docking solutions.[1–4] In the first applications of the ICM pseudo-Brownian docking we used explicit all-atom models for both proteins, and the correct arrangement was predicted solely on the basis of the global energy.[5,6] However, despite the improvements in computer hardware, this approach still remained too computationally costly for larger systems with unknown interface residues for both the receptor and the ligand. Thus, when confronted with the β-lactamase and its protein inhibitor,[7] we had to use the simplified models of proteins to speed up the rigid body search. Further acceleration came from the use of grid potentials to represent the bigger molecular partner. This hybrid approach was recently tested on a benchmark of 24 known uncomplexed structure pairs involved in protein–protein complexes.[8] The rigid body docking was followed by the refinement of the solutions and final scoring. This approach was later optimized and extended to unrestricted docking (Fernandez-Recio et al., forthcoming).

This optimized procedure, namely ICM Docking and Interface Side-Chain Optimization (ICM-DISCO), was used to generate automatically the predictions for the CAPRI experiment. In three of the seven proposed targets, our method found a good solution (defined as >50% of correct contacts) within the five submitted models.

MATERIALS AND METHODS

Energetics

The ICM-DISCO protein–protein-docking method is a variation of a previously described two-step docking procedure,[8] based on the ICM pseudo-Brownian minimization,[5] and Biased Probability Side-Chain Optimization[9] as implemented in the MolSoft ICM 2.8 program.[10] The interaction energy was represented by the following grid potentials,[11,12] precalculated on a grid[8,13] surrounding the 10 Å vicinity of the whole receptor: the truncated (maximum 1.0 kcal mol^{-1}) van der Waals potential (E_{vw}), an electrostatic potential corrected for the solvation effect[14] (E_{el}^{solv}), the hydrogen-bonding potential (E_{hb}), and a hydrophobicity potential (E_{hp}). The solvation energy (E_{solv}), based on atomic solvent-accessible surfaces,[15] was added to the total energy to revaluate the docking solutions obtained from unbound subunits (Eq. 1).

$$E = E_{vw(max\ 1.0)} + 2.16E_{el}^{solv} + 2.53E_{hb} + 4.35E_{hp} + 0.20E_{solv}$$

$$(1)$$

Rigid Body-Docking Simulations

To improve conformational sampling, multiple starting positions of the ligand around the receptor were defined as

---

*Correspondence to: Dr. Ruben Abagyan, Department of Molecular Biology, TPC-28 The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: abagyan@scripps.edu

TABLE I. Our Best models Submitted for CAPRI

Target	Model[a]	Correct contacts (%)	RMSD (Å)[b]	Comments
HPrK (*unbound*)-Hpr (*unbound*)	5	4	15.3	Correct solution not predicted because of large movement of receptor surface helix. Binding patch of the ligand correctly predicted in model 1
Vp6 (*unbound*)-Fab (*bound*)	3	4	26.6	Correct solution missed because of technical/human error due to multimolecular ligand (it ranked 5 after fixing it). Binding patch of both receptor and ligand correctly predicted in model 3.
Hemagglutinin (*unbound*)-Fab (*bound*)	2	71	4.0	OK
α-Amylase (*unbound*)-VHH (*bound*)	4	0	39.2	Rescoring made the correct answer (RMSD 0.2 Å) to rank 7. It ranked 1 after rigid docking (!)
α-Amylase (*unbound*)-VHH (*bound*)	3	11	17.7	Not found
α-Amylase (*unbound*)-VHH (*bound*)	1	72	1.5	OK
TCR-β (*unbound*)-SpeA (*unbound*)	5	59	3.4	OK

[a]Model number as ranked by our scoring function (Eq. 2; see Results for more details).
[b]RMSD of the corresponding model with respect to the X-ray structure, calculated for the C^α atoms of the ligand interface residues (defined as those ones with at least one atom within 4 Å distance from the receptor) when only the receptor molecules (model and X-ray structure) are superimposed. All RMSD values given in this work are calculated as described here.

follows. A series of points, with an average distance of 15 Å between neighbors, were generated around the receptor molecule. For the ligand, surface points were defined on a dodecahedron as previously described[8] for the first round of CAPRI (targets 1–3). However, for the second round (targets 4–7), ligand surface points were defined in the same way as for the receptor. Each combination of one receptor point (i_{Rec}), one ligand point (i_{Lig}) and one of six 60° rotations (i_{Rot}), was a starting point for a local stochastic search.

The sampling of rotational and translational degrees of freedom of the ligand starting from position {$i_{Rec}, i_{Lig}, i_{Rot}$} was performed by a pseudo-Brownian Monte Carlo minimization[5] implemented in the MolSoft ICM 2.8 program.[10] The calculation was parallelized on a Linux cluster, with all simulations for a given i_{Rec} assigned to a separate CPU. A single computational job took 2–7 h on a 667 MHz Alpha processor or 4–10 h on a 700 MHz Pentium III.

Side-Chain Refinement of Ligand-Binding Residues

In the absence of any experimental information, the 400 lowest energy solutions were selected, and this set was further compressed[16] to remove redundant conformations with RMSD for the interface C^α atoms of the ligand smaller than 4 Å. These solutions were further optimized by using fully flexible interface ligand side-chains and a grid map representation of the interaction energy as previously described.[8] The intermolecular energy was based on the same grid potentials used in the docking step, with the van der Waals truncated to 1.5 kcal mol^{-1}. During simulations, the internal conformational energy of the flexible parts of the ligand (E_{int}) was added, including internal van der Waals interactions (not truncated), hydrogen bonding, and torsion energy calculated with ECEPP/3 parameters,[17] and the coulomb electrostatic energy with a distance-dependent dielectric constant ($\epsilon = 4*r$).

The final scoring function used the intermolecular grid-based energy terms and the surface-based solvation energy[15] to revaluate the refined conformations (Eq. 2).

$$E = E_{vw(max\ 1.5)} + 2.16E_{el}^{solv} + 2.53E_{hb} + 4.35E_{bp} + 0.53E_{solv} \tag{2}$$

RESULTS
Target 1 (Unbound HPrK-Unbound HPr)

The whole HPrK hexamer (including phosphates) was used as a receptor, and the first subunit of the unbound dimeric HPr as a ligand. The 44 i_{Rec} points in the vicinity of the first receptor subunit out of a total of 80 around the hexamer were used to generate initial positions for rigid body docking. The 400 lowest energy rigid body solutions were compressed to a total of 258, which were selected for further refinement (see Materials and Methods). After refinement, a biological filtering step selected 74 solutions with H15 or D46 (expected phosphorylable ligand residues) within 15 Å of distance from the phosphate molecule. The five lowest energy nonsymmetrical solutions were submitted to CAPRI. The best of the submitted models (number 5) predicted only 4% of the native interatomic contacts in the complex X-ray structure[18] (Table I). A near-native solution (RMSD of the C^α atoms of the ligand interface residues when the receptor is superimposed onto the crystallographic structure: 4.8 Å) was found but not selected for refinement because of its very high energy value (rank 6824). Large unanticipated rearrangement of a receptor helix is the clear reason for the failure. It is of interest that 81% of the contact atoms of the ligand in the lowest energy model submitted were correct, which suggests the docking procedure recognized a general "protein-philic" patch. In fact, the same interface is also involved in crystallographic HPr homo-dimer formation (PDB 1sph).

Target 2 (Unbound VP6-Bound Fab)

In the rigid body-docking step, the structure of the unbound trimeric VP6 (excluding metal ions and water molecules) was used as a receptor, and the bound Fab structure as a ligand. The 40 i_{Rec} points in the vicinity of the first receptor subunit out of a total of 86 around the trimer were used to generate initial positions for rigid body docking. The resulting 9123 rigid body-docking solutions were filtered according to the number of contacts of the ligand (Fab) in the CDRs. We selected docking solutions with greater than four intermolecule atomic contacts in the CDR residues and less than two contacts in the constant regions. The 400 lowest energy conformations (out of the 738 that passed the filtering) were compressed to a total of 311, which were selected for further refinement.

Because the Fab coordinates were taken from the complex, this molecule was represented on the grid, so that the VP6 interface residues were refined. Then, we checked again the Fab contacts in the CDRs, using now a stricter criterion. For that, a contact ratio (number of CDR contacts divided by total number of contacts) was calculated for all the refined solutions. A total of 132 solutions passed this final filter (contact ratio $\geq$ 0.6). Symmetry-related solutions were removed, and the five lowest binding energy models were submitted to CAPRI. The best submitted model (number 3) predicted only 4% of the native interatomic contacts in the complex X-ray structure (Vaney and Rey, in preparation). Similar to target 1, in one of the submitted models (number 3), the receptor and ligand interfaces were correctly predicted (81% and 85% of correct contacts, respectively), although the mutual orientation of receptor and ligand was wrong. After analyzing the rigid body solutions, a near-native conformation (RMSD 7.3 Å) was found with rank 83. Later we discovered that this solution should have had rank 5 after refinement (with an RMSD of 6.1 Å), but it was not properly refined because of technical problems in the algorithm originated by the multimolecular ligand (these problems were fixed later, but we could not recalculate on time the solutions for target 2).

Target 3 (Unbound Hemagglutinin-Bound Fab)

In the rigid body-docking step, the structure of the unbound hexameric hemagglutinin (including carbohydrate molecules) was used as a receptor, and the bound Fab structure as a ligand. The 47 i_{Rec} points in the vicinity of the first two symmetry-independent receptor subunits out of a total of 114 around the hexamer were used to generate initial positions for rigid body docking. The resulting 10,937 docking solutions were filtered according to the number of contacts of the ligand (Fab) in the CDRs (as in target 2). The 400 lowest energy solutions (out of 769 that passed the filtering) were compressed to a total of 328 conformations, which were selected for further refinement.

The Fab was now represented on the grid, so that the hemagglutinin interface residues were refined. The contact ratio (CDR contacts/total contacts) was calculated for all the refined solutions, and a total of 115 solutions passed this final filter (contact ratio $\geq$ 0.6). Symmetry-related solutions were removed, and the five lowest energy conformations were submitted for the CAPRI. The best of our models (number 2) predicted 71% of the native interatomic contacts in the complex X-ray structure,[19] the highest number among the models submitted by all participants. As can be seen in Figure 1(a), the quality of this model is excellent (RMSD 4.0 Å), especially considering the size of the hemagglutinin surface and the fact that no restrictions of any kind were imposed for the binding to the Fab. The optimization of the ligand binding side-chains was essential to select the near-native solution (this model was ranked as 259th before refinement).

Target 4 (Unbound α-Amylase-Bound VHH)

In the rigid body-docking step, the structure of the unbound α-amylase was used as a receptor, and the bound VHH structure was used as a ligand. All 41 i_{Rec} points around the receptor and 16 i_{Lig} points around the ligand were used to generate initial positions for rigid body docking. The resulting 12,286 rigid body-docking solutions were filtered according to the number of contacts of the ligand (VHH) in the CDRs. The CDR contact ratio (CDR contacts/total contacts) was calculated for all the solutions, which were sorted by this value. The 400 solutions with the largest ratio values (minimum contact ratio = 0.58) were compressed to a total of 300 conformations, which were selected for further refinement.

The VHH was now represented on the grid, so that the α-amylase interface residues were refined. After sorting the 300 refined conformations by the scoring function (Eq. 2), the five lowest energy models were submitted to CAPRI. None of these models predicted a single native interatomic contact of the complex x-ray structure.[20] However, it is interesting that we actually found a very good solution with RMSD 0.2 Å (!) ranked 7 after refinement. What is even more striking is that this solution was ranked 1 (!) after rigid body docking.

Target 5 (Unbound α-Amylase-Bound VHH)

In the rigid body-docking step, the structure of the unbound α-amylase was used as a receptor, and the bound VHH structure was used as a ligand. A total of 41 i_{Rec} points around the receptor and 16 i_{Lig} points around the ligand were used to generate initial positions for rigid body docking. The resulting 12,061 conformations were filtered according to the CDR contact ratio for the VHH ligand (as calculated for target 4). The 400 solutions with the largest ratio values (minimum contact ratio = 0.72) were compressed to a total of 290 conformations, which were selected for further refinement.

The refinement was performed as described for target 4. The 290 refined solutions were sorted by the scoring function (Eq. 2), and the five lowest energy models were submitted to CAPRI. The best of our models (number 3) predicted only 11% of the native interatomic contacts described in the complex X-ray structure.[20]
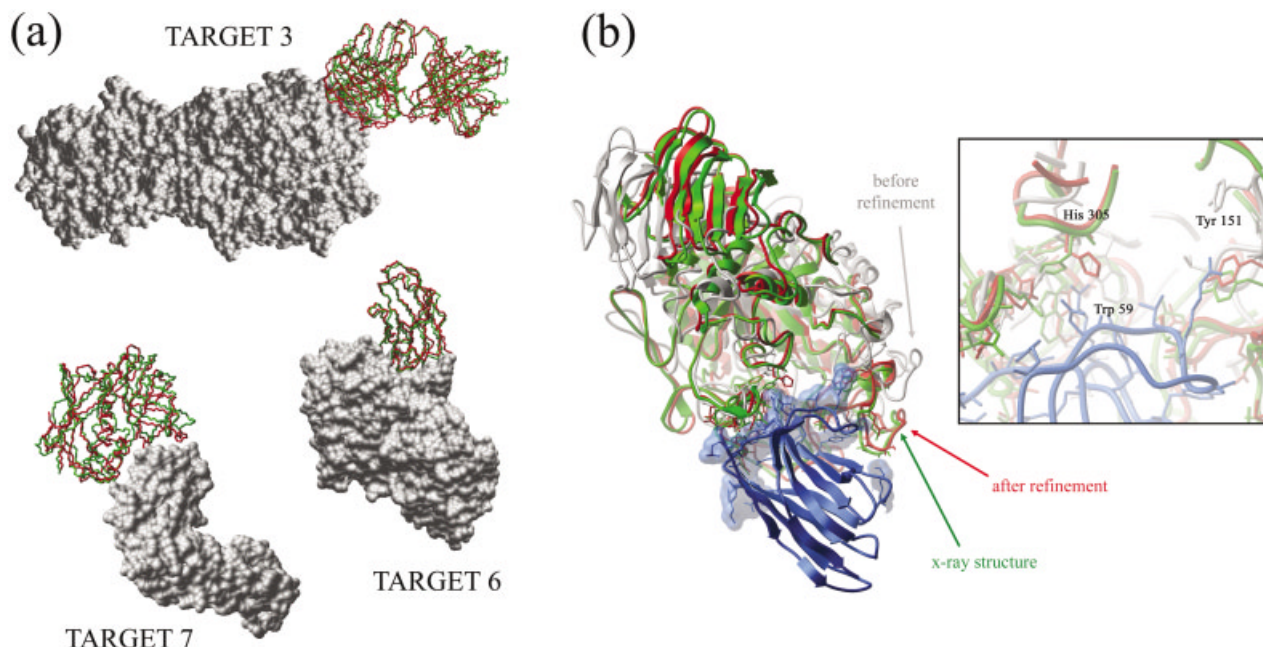
Fig. 1. **a:** Representation of our best submitted models for targets 3, 6, and 7. Ligand backbone is represented in red, and receptor molecular surface in white. The position of the ligand in the complex X-ray structure is represented in green (after superimposing the receptor molecules). **b:** Our best submission for target 6 (model 1, represented in red) is compared to the X-ray structure (in green), after superimposing the VHH molecules (in blue). The refinement step dramatically improved both the energy of the near-native solution (ranked 1 after final scoring of the refined solutions) and the geometry (the same solution before refinement is shown in white). It can be clearly seen in the inset that the refined interface (red) is much closer to the X-ray structure (green) than the unrefined rigid body solution (white).

Target 6 (Unbound α-Amylase-Bound VHH)

In the rigid body-docking step, the structure of the unbound α-amylase was used as a receptor, and the bound VHH structure was used as a ligand. All 41 i_{Rec} points around the receptor and 14 i_{Lig} points around the ligand were used to generate initial positions for rigid body docking. The resulting 12,754 conformations were filtered according to the CDR contact ratio for the VHH ligand (as calculated for targets 4 and 5). The 400 solutions with the largest ratio values (minimum contact ratio = 0.75) were compressed to a total of 285 conformations, which were selected for further refinement.

The refinement was performed as described for targets 4 and 5. The 285 refined solutions were sorted by the scoring function (Eq. 2), and the five lowest energy models were submitted to CAPRI. As can be seen in Figure 1(a), the best of our models (number 1) was very close to the reported X-ray structure of the complex,[20] predicting 72% of the correct interatomic contacts (RMSD 1.5 Å). The optimization of the ligand binding side-chains gave excellent results. As can be seen in Figure 1(b), not only did it help to rerank the near-native solution from the 44th position (before refinement) to the 1st one (after refinement) but also it improved the RMSD of the C^α atoms of the ligand-binding residues with respect to the corresponding ones in the X-ray structure (from 7.4 Å before refinement to 1.5 Å after refinement).

Target 7 (TCR-β-SpeA)

The unbound TCR-β was used as a receptor, and the unbound SpeA was used as a ligand. All 30 i_{Rec} points

around the receptor and 26 i_{Lig} points around the ligand were used to generate initial positions for rigid body docking. The resulting 8639 solutions were filtered according to the CDR contact ratio (CDRs contacts/total contacts), and a total of 568 conformations with contact ratio > 0.5 were selected. After removing solutions that would clash with the α-TCR (assuming the same TCR α/β interaction as in PDB 1tcr), the 400 lowest energy solutions were compressed to a total of 261 conformations, which were selected for further refinement. The four lowest energy solutions (sorted by Eq. 2) were submitted to CAPRI as models 1–4. For our fifth model, we compared all refined solutions with the known structure of the complex between TCR-β and Seb (52.8% sequence identity with SpeA, PDB 1sbb). We found that the conformation ranked 170 after refinement was close to this homologous structure, so we artificially included this solution as our model number 5. As can be seen in Figure 1(a), this submitted model was very close to the reported X-ray structure of the complex,[21] predicting 59% of the correct interatomic contacts (RMSD 3.4 Å).

DISCUSSION AND CONCLUSIONS

In all cases the rigid body-docking procedure found a near-native solution. In five of seven cases, this solution was ranked within 400 lowest energy conformations (after applying the CDR filter for antibodies). Furthermore, the interface refinement improved ranking and/or geometry. The most exciting result was a dramatic improvement of the near-native solution for target 6, for which the rank went from 44 to 1 and the RMSD from 7.4 Å to 1.5 Å (!), as

can be seen in Figure 1(b). Similarly, we observed improvements for targets 2 (once the problem with its refinement was fixed) and 3. However, in one case (target 4), the ranking was worse after the refinement, which seems to be due to rescoring rather than geometrical deterioration (RMSD improved from 0.6 Å to 0.2 Å after refinement).

Our results in this CAPRI blind test were in line with previous conclusions[8]: our models predicted >50% of the correct interatomic contacts in three of the seven targets [Table I; Figure 1(a)]. The CAPRI experiment has been an overall enriching experience that allowed us to check the applicability of our protein–protein docking procedure in a completely blind situation and helped us to formulate the next steps toward the solution of the protein docking problem.

REFERENCES

1. Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
2. Norel R, Petrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. Proteins 1999;36:307–317.
3. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.
4. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol 2002;12:28–35.
5. Abagyan R, Totrov M. ICM: a new method for structure modeling and design: applications to docking and structure prediction from the distorted native conformation. J Comp Chem 1994;15:488–506.
6. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. Nat Struct Biol 1994;1:259–263.
7. Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MN. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. Nat Struct Biol 1996;3:233–239.
8. Fernandez-Recio J, Totrov, M., Abagyan, R. Soft protein-protein docking in internal coordinates. Protein Sci 2002;11:280–291.
9. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994;235:983–1002.
10. ICM 2.8 Program manual. San Diego: MolSoft LLC; 2000.
11. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins 1997;Suppl 1:215–220.
12. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
13. Totrov M, Abagyan R. Protein-ligand docking as an energy optimization problem. In: Raffa RB, editor. Drug-receptor thermodynamics: introduction and applications. New York: John Wiley & Sons, Ltd.; 2001. p 603–624.
14. Fernandez-Recio J, Totrov M, Abagyan R. Screened charge electrostatic model in protein-protein docking simulations. Pac Symp Biocomp 2002;7:552–565.
15. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1:227–235.
16. Abagyan R, Argos P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 1992;225:519–532.
17. Nemethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HA. Energy parameters in polypeptides. 10. Improved geometric parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 1992;96:6472–6484.
18. Fieulaine S, Morera S, Poncet S, Mijakovic I, Galinier A, Janin J, Deutscher J, Nessler S. X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr. Proc Natl Acad Sci USA 2002;99:13437–13441.
19. Barbey-Martin C, Gigant B, Bizebard T, Calder LJ, Wharton SA, Skehel JJ, Knossow M. An antibody that prevents the hemagglutinin low pH fusogenic transition. Virology 2002;294:70–74.
20. Desmyter A, Spinelli S, Payan F, Lauwereys M, Wyns L, Muyldermans S, Cambillau C. Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. J Biol Chem 2002;277:23645–23650.
21. Sundberg EJ, Li H, Llera AS, McCormick JK, Tormo J, Schlievert PM, Karjalainen K, Mariuzza RA. Structures of two streptococcal superantigens bound to TCR beta chains reveal diversity in the architecture of T cell signaling complexes. Structure 2002;10:687–699.