SUPPORTING INFORMATION FOR WORLD WIDE WEB EDITION

Peptide structural prediction algorithm. Prediction algorithms can be divided into two parts, the scoring function and the conformational search. In this study, the chosen scoring function is among the most highly developed and well tested available. It is a physics-based second generation all-atom force field (OPLS-AA) with a GB/SA solvent model using software from Schrodinger, Inc. The scoring function evaluates the energy minimized structure and has been recently tested in the literature against databases of over 48,000 decoy structures from which it identified 30 out of 32 native structures of small proteins as the lowest energy (25). In the two cases where a decoy scored lower than the native structure, the decoy was only 0.6 nm RMSD from the native structure. Furthermore, there was a correlation between this scoring function and the RMSD of the decoys from the native protein structure (25). Although more complex than many other types of scoring functions, this one is among the most accurate.

Conformational searching methods include techniques based on distance geometry, libraries, energy minimization, low mode (normal mode), Monte Carlo, and dynamics and simulated annealing. Each method has a different efficiency at searching a particular characteristic of the conformational landscape. Consequently, the most rapid searching strategies often employ a combination of techniques. This study employs each of the major forms of conformational searching as implemented in the commercial software (Schrodinger, Inc.) to form a new algorithm to predict peptide structure from its sequence. The procedure as implemented for peptides with high helical character is described below:

1. The sequence is evaluated using conventional secondary structure prediction algorithms as implemented in ANTHEPROT software (http://antheprot-pbil.ibcp.fr). In both of the cases

presented here (TnT N47 and the villin headpiece, 1VII.pdb), some algorithms predicted helical content as high as 96%-97%.

2. Given the prediction of high helical content, the peptide is initially constructed as a helix from the fragment library in the Maestro (Schrodinger, Inc.) program.

3. A distance geometry method is implemented by running energy minimizations with harmonic restraints between hydrophobic side chain residues. In TnT N47, only 3 strongly hydrophobic residues are present, so all permutations of possible interactioning residues were evaluated. For the villin headpiece, millions of permutations were possible due to the greater number of hydrophobic residues, so 100 permutations were sampled. The distance constraints were set at 0.5 nm ± 0.2 nm between the tertiary carbons at the beta or gamma position of each side chain. Energy minimizations proceeded to convergence at a gradient of 0.05 kJ/(mol-Å) with a conjugate gradient method. Structures were eliminated if their energy was greater than 5 kcal/mol over the lowest energy from constrained minimization.

4. Each surviving structure was subjected to a stochastic (Langevin) dynamics simulation for 1 ns. The protocol involved releasing all constraints, energy minimization, heating to 300 °K, and dynamics for 1 ns. The SHAKE algorithm was applied to all bonds. The final structure was energy minimized.

5. A low mode search was then applied for 20 steps. Only the lowest energy structure was saved.

6. A mixed mode Monte Carlo and large scale low mode conformational search was performed for 1000 steps. Energy minimization to convergence at a gradient of 1 kJ/(mol-Å) was used

initially, then the lowest energy structures were minimized to convergence at a gradient of 0.05 kJ/(mol-Å). Default settings were used.

7. The lowest energy structure was used for a hybrid Monte Carlo and stochastic dynamics simulation at 300 °K for 100 ps with 1.5 fs steps and the SHAKE algorithm. The final structure was then seeded into a mixed mode Monte Carlo and low mode conformational search for 10 steps. The lowest energy structure of the set, including the structure before dynamics simulation, was retained.

8. Next a stochastic dynamics simulation was performed for 1 ns. The final strucure was seeded into a mixed mode Monte Carlo and low mode conformational search for 10 steps. The lowest energy structure of the set, including the structure before dynamics simulation, was retained.

9. Step 7 is repeated until no lower energy structures are found.

The results for the villin headpiece and TnT N47 are plotted in Figure 1. The data demonstrate that the above algorithm produces structures increasingly more similar to the native as evaluated by both the energy scoring function and RMSD to the average NMR structure for the villin headpiece. Notably the final energy is only 16 kcal/mol greater than the native structure and the RMSD is within 0.67 nm. The results are comparable to other state of the art prediction algorithms for peptides. TnT N47 shares many properties with the villin headpiece such as both have high propensities for helix formation, and both are about the same size (667 atoms and 595 atoms, respectively). The TnT N47 yields similar improvements in energy during the simulations, so a near native final prediction is expected as with the villin headpiece.

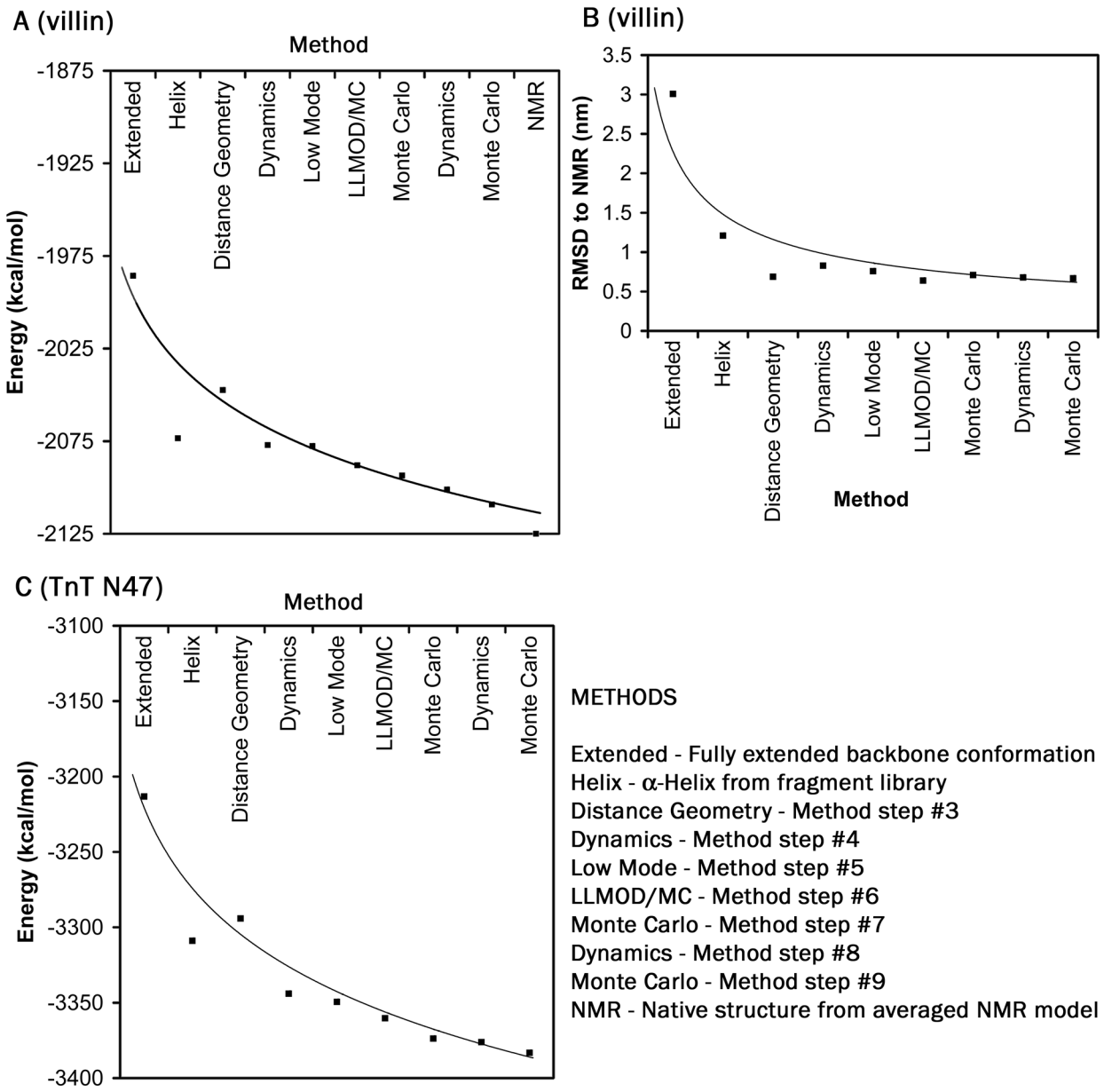

Figure 1. Profiles of energy and RMSD during the structure prediction algoritm. The profiles are for (A) relaxed energies of the villin headpiece, (B) RMSD to the native structure for the villin headpiece, and (C) relaxed energies for TnT N47. The methods are described in detail in the text. Note that the prediction for the villin headpiece is reduced to within 16 kcal/mol and 0.67 nm RMSD of the native NMR structure. Trends are drawn to guide the eye.

Binding free energy estimations. The interaction of calcium ions with TnT N47 was estimated in two ways. First the lowest energy structures for the TnT N47 with calcium bound and without calcium by using the conformational search techniques described above. The minimized energies of the TnT N47 with calcium bound less the energy of TnT N47 separated from a calcium ion by more than 10 nm so that no interaction occurs were used to give a rough approximation of the free energy of binding. A second method that is among the most accurate available for estimating absolute free energies of binding is the linear response method (37-39). This method is semi-empirical, since it relies on a scaling factor derived from experimental binding measurements.

To determine the scaling for the linear response method, a series of calcium chelates were selected that have stability constants approved by the National Institute of Standards and Technology (see Table 1 and Figure 2). Atomic models each of the chelates were constructed, and their global minimum energies were determined with and without calcium bound by conformational searching techniques. For all simulations in this work, the calcium ion parameters for the GB/SA solvent model were selected to yield a solvation energy of 381 kcal/mol to match its experimental value. Stochastic dynamics simulations were performed on each of the chelates in bound and unbound forms for 200 ps at 300 °K, and the average differences in electrostatic interactions (U_{ele}) and average differences in van der Waals interactions (U_{vdw}) were used to estimate the binding free energy (G_{bind}) with the equation:

$$G_{bind} = 0.5(U_{ele}) + \alpha(U_{vdw})$$

in which α is the scaling factor for the linear response method determined by fitting the binding free energies calculated from the known stability constants of the calcium chelates. The best fit value of α was determined to be 0.48. The scaling factor is highly force field dependent

and a literature value for α based on measurements in proteins with an OPLS-AA force field is 0.73 (37). Stochastic dynamics simulations were also performed on the bound and unbound forms of TnT N47, and the estimated binding free energies derived from the differences in average electrostatic and van der Waals terms are reported in Table 1. The error in the linear response method estimate for the binding free energy of the calcium chelates is 2 kcal /mol. By comparison, the error of the linear response method for ligands binding to proteins has been estimated at 4 kcal/mol (37). Within these errors the predicted model for TnT N47 binding calcium ion is in reasonable agreement with the amount of binding observed experimentally; however, a complete investigation of the binding energies of multiple calcium ions with the atomic model is not yet available.

Table 1. Experimental and Predicted Binding Free Energies (kcal/mol)[a]

Chelate	Experimental	Minimum	LRM α=0.73	LRM α=0.48
EGTA	-15.0	-29.8	-12.3	-14.7
EDTA	-14.7	-29.5	-14.2	-16.6
HEDTA	-11.2	-16.0	-6.8	-8.5
NTA	-8.9	-22.2	-8.5	-10.1
ATP	-5.2	-9.6	-3.7	-5.1
ADP	-3.9	-8.5	-1.8	-2.6
None	0	0	0	0
RMSD	0	10.6	2.4	1.6
TnT N47	-7	-6	-10	-9

[a]Minimum is the difference in global minimum energies. LRM is the linear response method.

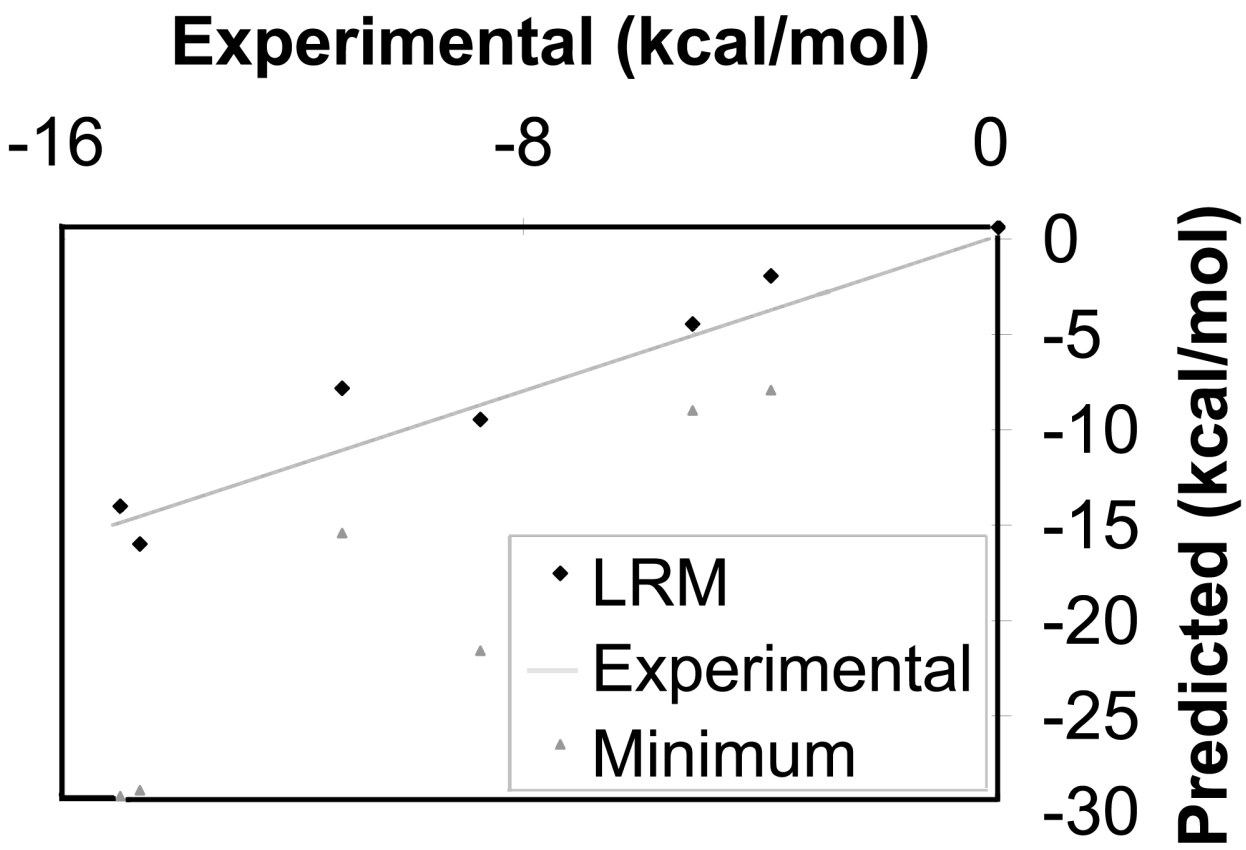

Figure 2. Experimental and predicted free energies of binding. The chelate data from Table 1 are plotted to visualize the correlation between predicted and experimental binding free energies. The diamonds show linear response method (LRM) data when $\alpha = 0.48$.The triangles indicate the difference in global minimum energies. The correlation coefficients are 0.97 and 0.96 for the diamonds and the triangles, respectively.

It is noted that the chelates display a broad range of structural changes upon binding calcium as documented in Table 2. NTA demonstrates the least amount of change upon calcium ion and could be described as a lock and key mechanism of binding. Curiously, EGTA is among the most calcium ion specific of the chelates, and it exhibits the greatest degree of structural change between the bound and unbound forms indicating an induced fit mechanism of binding. The

magnitude of structural change in some of the chelates is comparable to the degree of structural change in TnT N47 between its holo and apo forms.

Table 2. Diversity of structural changes in chelates upon binding calcium ion

Chelate	**EGTA**	**EDTA**	**ATP**	**HEDTA**	**ADP**	**NTA**
RMSD[a]	0.40 nm	0.28 nm	0.25 nm	0.23 nm	0.15 nm	0.06 nm

[a] The RMSD between the lowest energy bound and unbound forms is for all atoms except the ligand.

Preparation of TnT N47. TnT N47 is the NH_2-terminal 46 amino acids of chicken breast (fast) muscle TnT1 isoform (residues 2-47) and containing four of the metal-binding Tx motifs. cDNA encoding TnT N47 was constructed into an expression plasmid, expressed in *E. coli,* and purified by Zn^{2+}-affinity chromatography as previously demonstrated (*15*). Lyophilized TnT N47 was dissolved in 50 mM Tris, pH 8.0 buffer, containing 0.5 M KCl and 6 M urea. The solution was dialyzed against 50 mM Tris, 0.5 M NaCl, pH 7.8, and then 50 mM Tris, 100 mM NaCl, pH 7.8 buffer at 4°C, and subsequently ultracentrifugated at 100,000 *g* for 1 hr in a Beckman TL100 at 4°C. Then, G-25 gel filtration and Q-Sepharose ion exchange chromatography were performed to remove any trace amount of nucleic acid contaminants. Since the N47 peptide does not contain tryptophan, phenylalanine or tryrosine, its concentration was determined by UV absorbance at 205 nm and 215 nm by the equation (*20*):

Protein concentration (μg/ml) = 144 ($A_{215} - A_{225}$)

Purification of chicken breast muscle TnT. Adult chicken breast muscle TnT is predominately comprised of the isoforms containing the Tx cluster that bind to transition metal ions (*13*). Taking the advantage of metal binding properties, Zn^{2+}-affinity chromatography was performed to purify chicken breast muscle TnT as previously described (*14-15*). Chicken breast muscle was commercially purchased and freshly frozen. The breast muscle was thawed and homogenized with 0.1 M sodium phosphate, 2 mM imidazole, 1 M NaCl, 6 M urea, pH 7.0 buffer. The homogenate was subjected to $(NH_4)_2SO_4$ fractionation. The 45-60% saturation fraction was dissolved in 20 mM sodium phosphate pH 7.0 buffer, containing 2 mM imidazole, 0.5 M NaCl, 6 M urea and loaded on a Zn^{2+}-charged Chelating Sepharose Fast Flow column (Amersham Pharmacia), which is incorporated into a fast-purification liquid chromatography system (FPLC system, Pharmacia Biotech). Tx-positive TnT bound to the Zn^{2+} column was eluted with a 2-100 mM imidazole gradient. The isolated TnT peak was examined by 15%T, 2.5%C SDS-polyacrylamide gel electrophoresis (SDS-PAGE) to ensure the size and purity. Western blot using a monoclonal antibody (mAb) 3E4 (*16*) against a signature sequence APPP in the N-terminal region after Tx motifs of avian fast TnT was carried out to confirm the fidelity of purified TnT. TnT concentration was determined from molar extinction coefficient $\varepsilon_{280\ nm}$ of 26,000 $M^{-1}cm^{-1}$ in denature state, and 12,000 $M^{-1}cm^{-1}$ in native conformation (*17*). Purified TnT was stored in -70°C freezer.

Purification of rabbit skeletal muscle Tm. Rabbit skeletal muscle Tm was prepared from acetone powder by the method of Smillie *et al.* (*21*). Acetone powder was extracted with 1.0 M KCl, 0.5 mM dithiothreitol (DTT), pH 7.0 for overnight. Then, the extraction was subjected to isoelectric precipitation at pH 4.6, and followed by $(NH_4)_2SO_4$ fractionation. The 53%-65% saturation fraction was dissolved in 10 mM sodium phosphate, pH 7.0 buffer, containing 1 M KCl, 0.25 mM DTT, and 0.01% sodium azide and dialyzed against the same buffer at 4°C to remove remaining salt. Crude Tm after dialysis was loaded on the hydroxyapatite (Bio-Rad) column that incorporated with FPLC system. Tm was eluted with 10-200 mM phosphate gradient, and the fractions were examined by 15% SDS-PAGE in absence and presence of 5 M urea. The concentration of Tm was determined from $A_{280\ nm}$ = 0.24 $(mg/ml)^{-1}cm^{-1}$ (*22*) and a molecular weight of 33, 000 kDa (*23*).

Protein purity and fidelity. Figure 3A shows the SDS-PAGE gel of chicken breast muscle TnT from Zn^{2+}-chromatography. The Zn^{2+} affinity column is highly selective as shown by the result that TnT was the only major protein retained in the column after sample loading. Trace amounts of non-specific proteins were further removed from the column during wash (Figure 3A). Metal binding TnT was eluted from the column in high purity only when imidazole gradient was applied. Overloaded samples in Figure 3A not only prove greater than 99% purity of the TnT preparation as quantified by gel densitometry, but also provide evidence that potential contaminating calcium binding proteins such as TnC and parvalbumin, which have much lower molecular weights than TnT, have been eliminated. Hence, Ca^{2+} binding measured in the following experiments resulted exclusively from TnT. The western blot (Figure 3B) with mAb 3E4 (*16*) against the APPP epitope of avian pectoral TnT further demonstrated the fidelity of purified chicken breast muscle TnT.

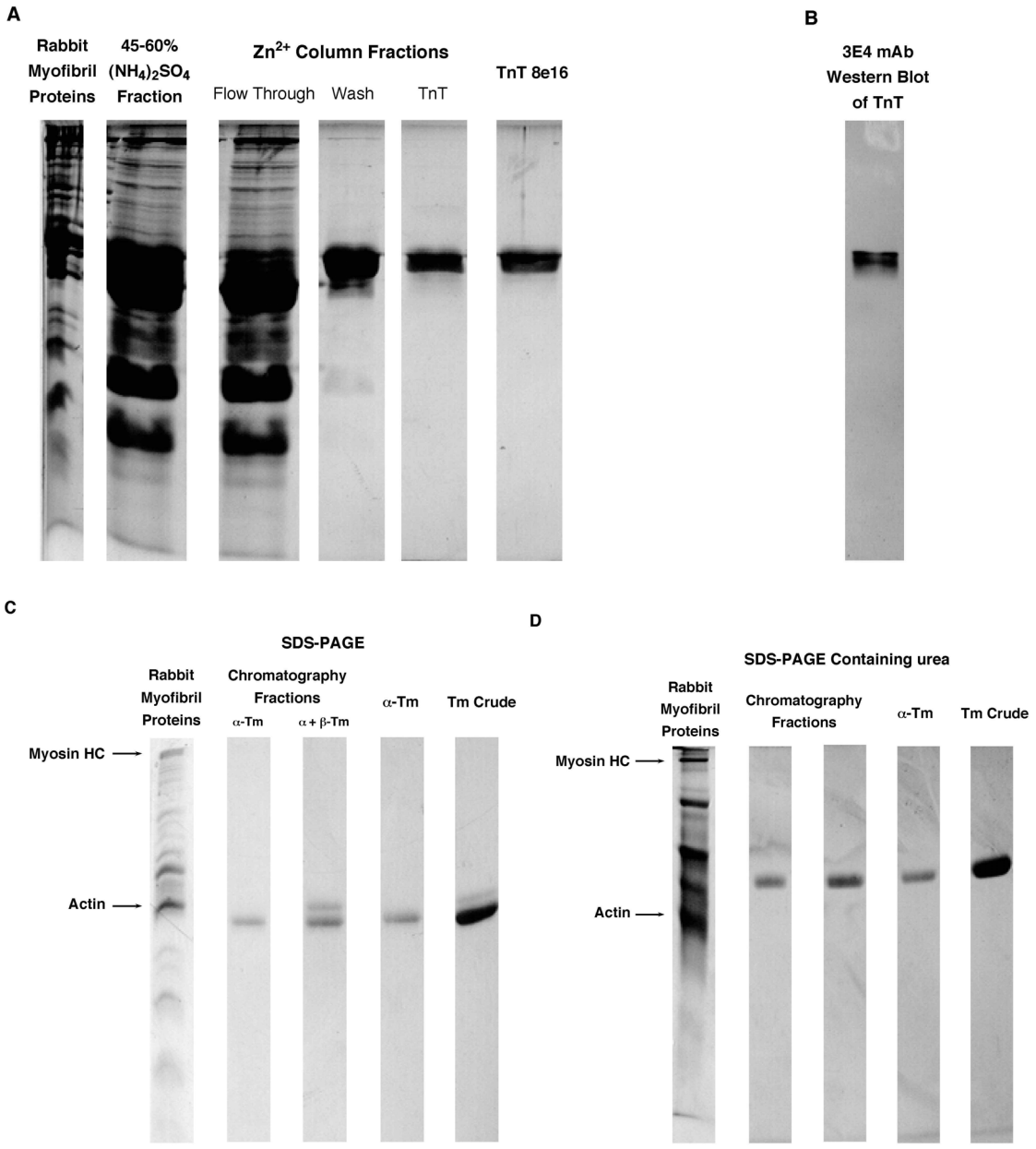

Figure 3. Purity and reconstitution of proteins assayed. Protein fractions from chromatography or dialysis were examined by 15% SDS-PAGE, and the gel picture was captured from a CCD camera. (A) TnT chicken breast muscle isoform was purified by Zn^{2+} affinity chromatography. Samples applied to each lane were as follows: rabbit myofibril proteins, chicken breat muscle protein extracts (the 45-60% $(NH_4)_2SO_4$ fraction), Zn^{2+} column fractions (flow through, wash, and

TnT eluted with imidazole gradient), and TnT 8e16 expressed from cDNA. (B) Western blot of TnT using mAb 3E4 proved the fidelity of TnT peak fraction. Rabbit skeletal muscle Tm was purified from acetone powder, and examined by SDS-PAGE in absence (C) and presence (D) of 5 M urea. Sample applied to each lane were as follows: rabbit myofibril proteins, Tm peak fractions (one contains α-Tm only, the other contains both α- and β-Tm), α-Tm purified from adult chicken ventricule as a standard (*21*), and Tm crude before column. Special property of Tm mobility shift in urea gel is clearly shown in (D).

Rabbit skeletal muscle Tm as purified from acetone powder by hydroxyapatite chromatography (*21*) was distinguished from TnT and actin by mobility shift in a urea gel during electrophoresis (*21*). In the urea gel (Figure 3D), it is clear that the protein bands shifted upward and appeared to have higher molecular weights than those in SDS polyacrylamide gel (Figure 3C). Moreover, α- and β-Tm were separated as shown in Figure 3C.