A Structural Model of the Erythrocyte Spectrin Heterodimer Initiation Site Determined Using Homology Modeling and Chemical Cross-linking*

Received for publication, August 21, 2007, and in revised form, October 26, 2007. Published, JBC Papers in Press, October 31, 2007, DOI 10.1074/jbc.M706981200

Donghai Li, Hsin-Yao Tang, and David W. Speicher.

From the Wistar Institute, Philadelphia, Pennsylvania 19104

Spectrin assembles into an anti-parallel heterodimeric flexible rod-like molecule through a multistep process initiated by a high affinity interaction between discrete complementary homologous motifs or “repeats” near the actin binding domain. Attempts to determine crystallographic structures of this critical dimer initiation complex have so far been unsuccessful. Therefore, in this study we determined the subunit-subunit docking interface and a plausible medium resolution structure of the heterodimer initiation site using homology modeling coupled with structural refinement based on experimentally determined distance constraints. Intramolecular and intermolecular cross-links formed by the “zero length” cross-linking reagent, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide were identified after trypsin digestion of cross-linked heterodimer complex using liquid chromatography-tandem mass spectrometry analysis. High confidence assignment of cross-linked peptides was facilitated by determination of cross-linked peptide masses with an uncertainty of a few parts per million using a high sensitivity linear ion trap mass spectrometer equipped with a Fourier-transform ion cyclotron resonance detector. Six interchain cross-links distinguished between alternative docking models, and these distance constraints, as well as three intrachain cross-links, were used to further refine an initial homology-based structure. The final model is consistent with all available physical data, including protease protection experiments, isothermal titration calorimetry analyses, and location of a common polymorphism, which affects the proportion of α chains from each allele that are incorporated into the red cell membrane by affecting heterodimer assembly. Hence, both heterodimer and tetramer assembly are critical steps in producing a functional membrane skeleton and disruption of one or both of these processes is frequently the cause of hereditary anemias.

The mechanism of spectrin heterodimer assembly is of particular interest because both subunits are primarily comprised of many tandem homologous “spectrin type” motifs or “repeats” ~106 residues in length. There are 20 repeats in the α subunit and 17 repeats in the β subunit (Fig. 1). All of these repeats are expected to have similar three-helix bundle conformations, and all repeats laterally associate with a homologous repeat in the complementary subunit. Despite the presence of many possible similar structures that could laterally dock with each other (Fig. 1) correct alignment and pairing of complementary repeats readily occurs. However, the structural features driving this assembly are only partially elucidated.

In earlier studies, we showed spectrin heterodimer assembly required a small region near the actin binding end of the molecule (β1–2 and α20–21) that initiated rapid, high affinity assembly (12, 13). Any spectrin fragments containing these sites could rapidly associate in the correct orientation with the complementary subunit, whereas fragments lacking these regions could not assemble. After dimer initiation, adjacent repeats rapidly laterally paired with an inter-chain partner, although these associations were very weak and were not detected in solution unless the dimer nucleation site was
Spectrin Heterodimer Initiation Site Model

![Diagram](Image)

**FIGURE 1.** Model of the human erythrocyte spectrin anti-parallel heterodimer. The structural domains of the anti-parallel α and β subunits are schematically represented as follows: numbered rectangles, homologous “spectrin type” repeats; loop, labeled as repeat 10 is actually a src SH-3 homology domain; diamonds, EF hands; large rectangle, actin binding domain. The “αα” represents a partial repeat at the N terminus of α spectrin involved in forming head-to-head tetramers as well as the closed hairpin form of the dimer (shown here). The “tail” at the C-terminal end of β spectrin is the non-homologous phosphorylated region.

covalently linked to the weaker repeats (12). This series of studies resulted in a “zipper” model of spectrin heterodimer assembly, which consisted of three discrete steps: 1) rapid, high affinity binding of the dimer initiation site; 2) rapid, low affinity lateral association of complementary repeats along the length of the monomer; and 3) latching the zipper by forming a closed dimer through a slow, moderate affinity, temperature-dependent association (14). We subsequently showed dimer initiation was weakened in high ionic strength buffers. In the same study homology modeling of the α20 and β2 repeats suggested their AB faces had complementary electrostatic surfaces that could drive initial docking of correct repeats through electrostatic interactions (15). Unfortunately, our repeated attempts to obtain a crystallographic structure of this complex have so far been unsuccessful, which is not surprising because crystallization of even single chain spectrin repeats has proven to be very challenging, presumably due to high molecular flexibility.

An alternative approach to structure determination is to use chemical cross-linking combined with identification of specific cross-linked sites using tandem mass spectrometry (MS/MS)2 (16, 17). Although chemical cross-linking is a very old technique, it has historically been very challenging to identify cross-linked peptides, because they are usually substoichiometric and difficult to separate, detect, and identify (17). This situation has improved substantially over the past several years as high sensitivity, high speed mass spectrometers capable of very precise mass determinations have become available. Although only a few cross-links are likely to occur under mild cross-linking conditions, these distance constraints can be of great value in distinguishing between alternative docking orientations as well as to test and refine structures based on homology modeling (16).

In this study, the docking interfaces of the red cell spectrin heterodimer initiation site region were experimentally determined using a “zero length” cross-linker coupled with analysis of cross-linked tryptic peptides using liquid chromatography-MAS/MS analysis on an LTQ FT-ICR mass spectrometer. Distance constraints from both intrachain and interchain cross-links were also used to further refine an initial homology-based structure. The final model correlates well with available biochemical and thermodynamic data, and it supports the hypothesis that initial alignment and pairing of correct α and β repeats from among many similar repeats in both subunits is driven primarily by long range electrostatic interactions.

**EXPERIMENTAL PROCEDURES**

**Construction of Expression Plasmids**

The following recombinant peptides were used in this study: α20–21 (repeats 20 and 21 of the human red cell spectrin α subunit, residues 2033–2259) and β1–2 (repeats 1 and 2 of the β spectrin subunit; residues 293–528). The repeats described are structural domains that use the phasing of Yan et al. (18). The design and construction of the α20–21 and β1–2 pGEX-2T expression plasmids have been previously described (13, 15).

**Expression and Purification of Recombinant Erythrocyte Spectrin Peptides**

The glutathione S-transferase-spectrin fusion proteins were expressed and purified as described (13, 15), except that β1–2 was expressed at 30 °C, α20–21 was expressed at 18 °C, and both of the proteins were purified from the soluble fraction after cell lysis. After initial purification on a glutathione-Sepharose column, peptides were cleaved from the glutathione S-transferase moiety using bovine thrombin (Sigma) at 37 °C for 3 h. NaCl was added to decrease the formation of secondary cleavage products at a final concentration of 0.5 M for β1–2. The cleaved spectrin recombinants were purified by rechromatography on a glutathione-Sepharose column. The α20–21 was then separated from residual glutathione S-transferase by anion-exchange chromatography on a 1-mL HiTrap-Q column (GE Healthcare). The α20–21 was bound to the column in Buffer A (65 mM NaCl, 5 mM sodium phosphate, 2.5 mM EDTA, 0.02, 0.1, 0.2, or 1 M EDC in addition to 0.5 M sulfo-NHS, was freshly prepared aqueous cross-linker solution, containing 0.02, 0.1, 0.2, or 1 M EDC in addition to 0.5 M sulfo-NHS, was added to 990 μl of α20–21/β1–2 complex (2 μM). Thus, molar excesses of 100, 500, 1000, and 5000 of EDC over the complex concentration were evaluated. For a control, 10 μl of water was added, instead of the cross-linker, to the complex solution. The reaction mixtures were incubated at room temperature, and

---

2 The abbreviations used are: MS/MS, tandem mass spectrometry; EDC, 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride; FT-ICR MS, Fourier transform ion cyclotron resonance mass spectrometry; HPLC, high performance liquid chromatography; sulfo-NHS, N-hydroxysulfosuccinimide; GPMAW, General Protein Mass Analysis for Windows; ASAP, Automatic Spectrum Assignment Program; MALDI MS, matrix-assisted laser desorption/ionization mass spectrometry; DOPE, Discrete Optimized Protein Energy.
200-μl aliquots were taken after 5, 15, 30, 60, and 120 min. The reactions were quenched by adding 20 μl of 110 mM aqueous hydroxylamine solution to each aliquot (final concentration, 10 mM). Before SDS-PAGE, the solutions were dialyzed against 25 mM NH₄HCO₃ overnight and lyophilized.

**SDS-PAGE and Trypsin Digestion**

Following separation of the reaction mixtures by one-dimensional SDS-PAGE, the bands of interest were excised and digested in gel as described previously using modified trypsin (Promega, Madison, WI) (19).

**Ion Trap FT-ICR MS**

Liquid chromatography-MS/MS experiments were performed on a hybrid ion trap FT-ICR mass spectrometer (LTQ-FT, Thermo Electron, San Jose, CA), equipped with a NanoLC pump and autosampler (Eksigent Technologies, Livermore, CA). Specifically, tryptic peptides were separated by reversed-phase HPLC on a 75-μm inner diameter × 15-cm PicoFrit (New Objective, Woburn, MA) nanocapillary column packed with 5 μm of MAGIC C18 AQ resin (Michrom BioResources, Auburn, CA). Solvent A was 0.58% acetic acid in Milli-Q water, and solvent B was 0.58% acetic acid in acetonitrile. Peptides were eluted at 200 nl/min using an acetonitrile gradient consisting of 1–50% B over 15 min, 50–80% B over 5 min, 80% B for 10 min before returning to 1% B in 1 min. The LTQ-FT mass spectrometer was set to perform a full MS scan (m/z 375–2000) in the FT-ICR with resolution at 400 m/z set to 100,000. Simultaneously, the six most intense ions exceeding a minimum threshold of 1000 were selected for MS/MS in the linear trap. The normalized collision energy was set to 30%, and ions subjected to MS/MS were excluded from repeated analysis for 30s. Singly charged ions were not subjected to MS/MS.

**Identification of Cross-linked Peptides**

Cross-linked peptides were identified using a combination of GPMAW (General Protein Mass Analysis for Windows) software, version 7.10 (Lighthouse Data, Odense, Denmark), SEQUEST Browser software (ThermoFisher Scientific), and the ASAP (Automatic Spectrum Assignment Program) software packages (16). MS/MS spectra of cross-linked peptide candidates were further analyzed by MS2Assign (16) and ProteinXXX, which is the cross-linking feature of GPMAW v.7.10.

**Construction of Homology Models of Laterally Associated α20–21 and β1–2**

**Search for Homology and Sequence Alignment**—The most suitable template for modeling the α20–21/β1–2 heterodimer was found via SWISS-MODEL Workspace tools (Template Identification) (20) and MODELLER (21). Sequence alignments were produced with ClustalW-1.83 (22), and sequence-structure alignments were done by MODELLER alignment2d.

**Homology Modeling**—The final sequence alignment was submitted to MODELLER 8v2 for generating a homology model of α20–21 and β1–2. Molecular graphics were illustrated using PyMOL (23). The dimer interaction was analyzed using the Protein-Protein Interaction Server (24).

**RESULTS**

**Chemical Cross-linking Reaction**—Monomers of the recombinant proteins were verified by MALDI MS, and potential irreversible aggregation of purified samples was measured shortly before the cross-linking reaction using analytical HPLC gel filtration (data not shown). These results together with prior characterization of these recombinants using CD, sedimentation equilibrium, and differential scanning calorimetry provided assurance these recombinant proteins were properly folded prior to cross-linking experiments.

The α20–21/β1–2 heterodimer was purified using HPLC gel filtration (Fig. 2A), and cross-linked using “zero length” cross-linking with EDC/sulfo-NHS, which cross-links acid groups (Asp, Glu, and the C terminus) to amines (ε-amino group of Lys and the N terminus) by creating an amide bond with elimination of an H₂O molecule. Both intra- and interchain cross-links are expected, because these two proteins contain high concentrations of dispersed glutamic acid, aspartic acid, and lysine residues. After the cross-linking reaction, the reaction mixtures were separated by one-dimensional SDS-PAGE, and the gels were stained with Coomassie Brilliant Blue. A range of cross-linking reaction conditions were evaluated as described under “Experimental Procedures,” and the optimal reaction condition was determined to be a 60-min reaction using a 1:5000 molar excess of EDC, because this produced a substantial amount of a cross-linked dimer complex without significant evidence of aggregation (Fig. 2B). The cross-linked α20–21/β1–2 complex is expected to migrate on SDS gels at ~55,000 Da depending on the extent of chemical cross-linking. Aggregation of proteins caused by excessive cross-linking was not observed for the cross-linked product under the 1:5000 molar ratio condition as evidenced by the absence of gel bands above the cross-linked dimer or at the top of the gel. Following SDS-PAGE separation of the cross-linked reaction mixtures,
the cross-linked complex was excised from the gel, subjected to in-gel digestion with trypsin, and analyzed on a hybrid ion trap-FT-ICR mass spectrometer.

Characterization of Inter- and Intramolecular Cross-linked Peptides Using Ion Trap FT-ICR MS—Non-cross-linked tryptic peptides from α20–21 and β1–2 were identified using the SEQUEST browser, and all MS/MS spectra corresponding to high confidence assignments from this analysis were removed from the dataset. The experimentally obtained monoisotopic masses for all remaining MS/MS spectra were compared with calculated masses of predicted tryptic peptides and theoretical cross-linked peptides using GPMAW 7.10 and ASAP (16). An initial mass error cut-off of 5 ppm was employed for this comparison.

MS/MS spectrum assignments of putative cross-linked peptides were performed with the aid of MS2Assign and the ProteinXXX program as well as manual de novo sequencing. Analysis of all spectra representing potential cross-linked peptides yielded high quality, high confidence matches for nine different masses arising from six specific intermolecular cross-links as summarized in Table 1. Fig. 3 shows the correlation of two representative MS/MS spectra with the assigned sequences, and assignments of observed ions are summarized in Tables 2 and 3. The nomenclature of cross-linked peptide fragment ions follows that proposed by Schilling et al. (26). The larger peptide is indicated by a subscript “A” and the smaller peptide by a subscript “B.” In the instrument used here, MS/MS spectra primarily result from random fragmentations at peptide bonds, and by convention, ions from the N-terminal region of a peptide are b-ions while the complementary C-terminal fragments are y-ions. In most cases, b- and y-ions can arise from either peptide, and some ions will include an intact cross-linked peptide. For example, ion 28 in Fig. 3A and Table 2 is designated as \((\text{A} + \text{B})^+\), which indicates it is a doubly charged b3 ion from peptide B that has an intact peptide A cross-linked to it via the lysine side chain at its N-terminal (see sequence diagram inset in Fig. 3A).

Sequence Alignment and Model of Laterally Associating α and β Repeats in the Dimer Initiation Site—An accurate sequence alignment is of utmost importance in building a useful three-dimensional homology model. The crystal struc-

Spectrin Heterodimer Initiation Site Model

### TABLE 1

| Exp. mass | m/z     | Error | Charge | Peptide locations | Cross-linked residues | Distance* | Å |
|-----------|---------|-------|--------|------------------|-----------------------|-----------|---|
| 501.7755  | 4.01    | 2     | α1–5/β234–238 | α1–β238               | 4.86                  |           |   |
| 905.1132  | 0.71    | 4     | α33–50/β140–144 | α39–β140              | 4.84                  |           |   |
| 679.0865  | 1.55    | 4     | α33–50/β140–144 | α39–β140              | 4.84                  |           |   |
| 915.7754  | 0.48    | 5     | α33–50/β140–144 | α39–β140              | 4.84                  |           |   |
| 549.8666  | 3.61    | 5     | α33–50/β140–144 | α39–β140              | 4.84                  |           |   |
| 951.4918  | 1.80    | 2     | α159–171/β1–12 | α164–β1               | 4.29                  |           |   |

Intramolecular cross-links

| m/z     | ppm | Å |
|---------|-----|---|
| 584.9731 | 3   | 7.98 |
| 500.2770 | 0.25 | 3.67 |
| 666.7009 | 0.20 | 3.67 |
| 812.7333 | 0.57 | 7.77 |
| 900.4513 | 0.35 | 3.25 |

*Distance between cross-linked atoms calculated using the Docking C model. Note that analyses of cross-link in proteins with known crystal structures, residues up to 10 Å apart in the crystal structure were shown to be cross-linked by EDC due to length and flexibility of the side chains being cross-linked (43).

### FIGURE 3

MS/MS spectra of representative intramolecular and intermolecular cross-linked peptides. A, cross-linked peptides ALE\(^{160}\)DLAQ\(\)LEK and K\(^{140}\)DNILR with precursor m/z 500.2770 and a charge state of +2. See Table 2 for the MS/MS assignments; B, cross-linked peptides MEENLSE\(^{39}\)PVHCVSLNEIR and AAMIR with precursor m/z 679.0865 and a charge state of +4. See Table 3 for the MS/MS assignments. Most ion assignments were from MS2Assign and ProteinXXX. The superscript numbers in the peptide sequences represent the cross-linked amino acid residue positions. The locations of the peptide bonds fragmented to produce the numbered ions are indicated on the sequences of the cross-linked peptides.


Table 2: MS/MS assignments for an intra-chain cross-linked peptide complex with the precursor ion [M+4H]^{4+} = 500.2770

| No. Peaks | Assignments | No. Peaks | Assignments |
|-----------|-------------|-----------|-------------|
| 1  | 147.16      | y_1A      | 16  | 518.31   | y_4A      |
| 2  | 175.15      | y_4A      | 17  | 527.53   | (b3)_Bint |
| 3  | 201.05      | (y_3B)^{-2} | 18  | 533.50   | (A_2)_Bint |
| 4  | 258.19      | (y_4B)^{-2} | 19  | 571.11   | (b4)_Bint |
| 5  | 276.27      | y_2A      | 20  | 585.11   | (b4)_Bint |
| 6  | 286.25      | y_2A      | 21  | 630.39   | y_5N      |
| 7  | 323.69      | (y_3A)^{-2} | 22  | 641.21   | (b5)_Bint |
| 8  | 359.18      | (y_3A)^{-2} | 23  | 646.34   | y_5A      |
| 9  | 389.28      | y_3A      | 24  | 676.79   | (b6)_Bint |
| 10 | 390.36      | (b4)_Bint | 25  | 717.39   | y_6A      |
| 11 | 401.43      | y_3A      | 26  | 742.37   | (A_2)_b2 |
| 12 | 428.03      | (b5)_Bint | 27  | 790.64   | (A_2B)_b3 |
| 13 | 451.72      | (b5)_Bint | 28  | 799.62   | (A_2B)_b3 |
| 14 | 491.77      | (b4)_Bint | 29  | 830.49   | (b10)_y2 |
| 15 | 515.28      | y_4A      | 30  | 945.43   | y_8A      |

Table 3: MS/MS assignments for an inter-chain cross-linked peptide complex with the precursor ion [M+4H]^{4+} = 679.0865

| No. Peaks | Assignments | No. Peaks | Assignments |
|-----------|-------------|-----------|-------------|
| 1  | 251.10      | b2_A      | 19  | 696.11   | (b7A)^{-2} |
| 2  | 280.20      | y_2A      | 20  | 701.11   | (y_13B)_Bint |
| 3  | 306.31      | y_2A      | 21  | 704.17   | b6_A      |
| 4  | 377.25      | y_2A      | 22  | 728.29   | (b14)_Bint |
| 5  | 390.46      | b3_A      | 23  | 731.43   | y_6A      |
| 6  | 417.44      | y_3A      | 24  | 737.67   | (y_14B)_A |
| 7  | 448.37      | y_3A      | 25  | 756.40   | (A_1B)_b2 |
| 8  | 495.80      | (y_3B)^{-2} | 26  | 766.58   | (b15)_Bint |
| 9  | 504.36      | b4_A      | 27  | 775.78   | (y_15B)_A |
| 10 | 531.32      | y_3A      | 28  | 779.62   | (A_1B)_b2 |
| 11 | 564.46      | (y_3B)^{-2} | 29  | 803.73   | (A_1B)_b2 |
| 12 | 575.67      | (b10)_Bint | 30  | 809.31   | (b10)_Bint |
| 13 | 617.19      | b2_A      | 31  | 818.64   | (y_16B)_A |
| 14 | 628.79      | (b11)_Bint | 32  | 830.51   | y_5A      |
| 15 | 644.46      | y_5A      | 33  | 847.51   | (A_2B)_b3 |
| 16 | 661.09      | (b12)_Bint | 34  | 861.83   | (y_17B)_A |
| 17 | 662.35      | (y_11B)^{-2} | 35  | 862.09   | (b10)_y2 |
| 18 | 670.52      | (y_12B)^{-2} | 36  | 990.49   | y_8A      |


ture of human skeletal muscle α-actinin 2 (PDB ID: 1HCl) with 28% sequence identity to α20–21 and 42% sequence identity to β1–2, was the best template for building homology models. Sequence alignments were produced with ClustalW-1.83, and MODELLER alignment.align2d, which is preferred for aligning a sequence with structure(s) in comparative modeling, was used to determine sequence-structure alignment (Fig. 4A). An additional consideration when modeling a non-covalent complex is how the two structures dock with each other. For our initial homology model of the α20–21/β2–2 complex we aligned α20–21 to the α-actinin repeats 3 and 4 (R3–4) and β1–2 with α-actinin repeats 1 and 2 (R1–2) using the α-actinin docking interface from its dimer crystal structure. This alignment and docking mapped each spectrin heterodimer repeat to its most homologous repeat in the α-actinin anti-parallel homodimer (Docking A in Fig. 4B). Based upon homology alone, this docking interaction seemed to be the most reasonable. However, this docking interface, which contains CA/AB and BC/BC interaction surfaces (Fig. 4B, Docking A), did not agree with our previous protease protection analyses of the heterodimer initiation site complex, which indicated AB faces were protected in the heterodimer interface region (15).

To model alternative heterodimer docking interfaces, the α20–21 and β1–2 monomers were docked using the ClusPro server (27). When using the ClusPro program, the user has the option of selecting DOT (28) or ZDOCK (29, 30) to perform rigid body docking, both of which are based on fast Fourier transform collision techniques. Applied to a benchmark set of 2000 conformations, the algorithm predicts at least one experimentally relevant complex structure within the top 20 predictions, and in ~30% of the cases, the best prediction is ranked first. Ten docking models were obtained using DOT and ZDOCK, respectively (data not shown). The returned models contained three kinds of inter-chain interactions: parallel interaction, “X”-like interaction, and anti-parallel side-by-side interaction. Because many previous experiments had shown the α and β subunits associate anti-parallel and side-by-side along the length of both chains to form a long flexible rod-like heterodimer, the first two types of interactions were discarded as unrealistic predictions. The third group of models had the correct anti-parallel orientation of subunits, but the predicted docking faces were inconsistent with the protease protection data (15). Therefore, none of the models predicted by ClusPro fit available experimental data.

We therefore reconsidered all published crystallographic structures of spectrin-type repeats. Of those available structures, only α-actinin R1–4 (PDB ID: 1HCI) and α-actinin R2–3 (PDB ID: 1QUU) form a physiologically relevant homodimer. Chicken brain spectrin α15–17 (PDB ID: 1U4Q) crystallized as an anti-parallel homodimer, however the interchain repeats are not aligned in register, as is expected for the α20–21/β1–2 dimer initiation site (15). Hence, we returned to using α-actinin as a docking template. The crystal structure of the α-actinin dimer contains three possible anti-parallel dimer docking orientations for two tandem repeats as illustrated in Fig. 4B. Development of the model using Docking A was described above. Models using Dockings B and C were obtained by first aligning α20–21 to R3–4 and β1–2 with R1–2 to build homology models of the spectrin monomers as described above. Alternative alignments were regarded as less favorable due to both the substantially higher homologies of the above pairings and differing lengths of repeats and locations of gaps/insertion. Alternative models of the spectrin dimer were then obtained by superimposing the monomer models onto the crystal structure of the human skeletal muscle α-actinin as illustrated in Fig. 4B (Docking models B and C). Modeling and optimization were performed with MODELLER.

Fig. 5 shows locations of observed cross-linked amino acid residues in models using the three alternative docking orientations. In the models using Docking A and B, cross-linked residues were often on opposite faces of the putative dimer (Fig. 5, A and B), and these distance constraints could not be used to refine the structures without extensive distortion of the prototypical three helix bundle observed in all crystal structures of spectrin type repeats determined to date. In contrast, the Docking C model is consistent with these cross-linked data (Fig. 5C), and the distance between the two residues in each identified cross-link (see Table 1) could be minimized during model refinement without excessively distorting the basic three helix bundle structures of these repeats. The quality of the Docking C-based structure was subsequently tested by PROCHECK (31). Ramachandran maps of the model revealed that the α20–
21/β1–2 complex contained 96.4% of non-Gly-non-Pro residues in most favored, 3.4% in additionally allowed, 0.2% in generously allowed, and 0.0% in disallowed regions. The final homology model of α20–21/β1–2 heterodimer also passed Whatcheck (20) and MODELLER DOPE (Discrete Optimized Protein Energy) (21) (data not shown). These evaluation data together with the physical constraints dictated by the cross-links indicate a high quality model with correct docking of the subunits.

In the modeled complex (Docking C model), the interacting surface on α20–21 consists of the AB face of α20 and the AC face of α21, which interact with the AC face of β2 and the AB face of β1, respectively. The interface surface along the long axis of the molecule buries 12.6% (1929 Å²) of the accessible
surface area of \( \alpha\text{20–21} \) and 12.8% (1841 Å²) of the accessible surface area of \( \beta\text{1–2} \) (determined with a probe radius of 1.4 Å). The interface involves 46% polar and 54% nonpolar atoms for \( \alpha\text{20–21} \), and 43% polar and 57% nonpolar atoms for \( \beta\text{1–2} \). The hydrophobic effect is not a dominant driving force in the interaction of \( \alpha\text{20–21} \) with \( \beta\text{1–2} \) because no significant hydrophobic patches were located in complementary positions on the protein interaction surfaces as indicated by LIGPLOT (32) analysis of the refined Docking C model. In contrast, the model reveals 10 intermolecular hydrogen bonds (H-bond) (3.9 Å cut-off) and 10 putative salt bridges (Table 4). The negatively charged residues are primarily contributed by \( \alpha\text{20–21} \), and the positively charged residues are primarily contributed by \( \beta\text{1–2} \) for these electrostatic interactions (Fig. 6A and Table 4). These complementary electrostatic interactions are very likely to drive recognition and pairing of the correct \( \alpha \) and \( \beta \) repeats during heterodimer assembly as we previously proposed (15).

Another interesting comparison is to evaluate the location of the six residue deletion in the \( \alpha\text{21} \) repeat associated with the \( \alpha^{\text{LLEY}} \) polymorphism in the alternative docking models. This polymorphism is common in the human population, and it leads to low incorporation into the red cell membrane of the protein from the allele carrying this mutation. The polymorphism involves three linked mutations, which include: a point mutation at codon 1857 that changes a Leu to a Val, a point mutation in intron 45, and a point mutation in intron 46. Presumably as a result of one or both intron mutations, there is a 50% skipping of incorporation of exon 46 (residues 2177–2182) in the expressed protein (10). We previously showed the Leu → Val substitution did not affect incorporation of \( \alpha \) subunits into dimers and hence into red cell membranes, but the deletion of the 6 amino acids encoded by exon 46 prevented these subunits from being incorporated into the membrane, by reducing dimer binding affinity (33). These six amino acids are located in the \( \alpha\text{21} \) A helix, which is in the dimer contact site in our Docking C model (Fig. 6B). Similarly, the \( \alpha\text{21} \) A helix is in the dimer interface in the B model (see Figs. 4B and 6B). In contrast, in model A, these residues are not in or near the dimer interface region (Fig. 6B). Hence, both Models B and C, but not model A, are consistent with the observed abolishment of heterodimer assembly by the exon 46 deletion. Although the location of the \( \alpha^{\text{LLEY}} \) polymorphism is consistent with models B and C, Docking model B is not consistent with the cross-linking data, so the Docking C model involves the best docking orientation of the spectrin heterodimer initiation site.

### DISCUSSION

The structural features that determine correct docking of the specialized spectrin repeats in the dimer initiation site were previously ambiguous. Three feasible models of the interchain contact surfaces were suggested by the crystal structure of the \( \alpha \)-actinin homodimer (34). To further characterize this critical spectrin self-assembly site and to distinguish the most feasible inter-subunit docking surfaces, we exploited the improved
capacity of high mass accuracy hybrid ion trap mass spectrometers to identify interchain cross-links in complex protease digests.

The use of chemical cross-linking to either identify binding partners or to obtain low to moderate resolution insights into protein conformation is a conceptually simple, classic strategy (16), but the purification and identification of cross-linked peptides using conventional protein chemical methods requires large amounts of sample and is usually very challenging. Cross-linking reactions are frequently very incomplete and attempts to drive them to completion increases the likelihood that native conformation will be disrupted and artifactual cross-links might be identified. In addition, the sub-stoichiometric cross-linked peptides are often especially difficult to purify and characterize. The application of mass spectrometry to this problem has greatly improved the feasibility of obtaining useful information on small amounts of cross-linked proteins, and recently the combination of high resolution FT-ICR MS with cross-linking has been demonstrated to be a fairly robust strategy for rapidly defining interfaces between proteins (35).

So that distance constraints would be as specific as possible, we used the “zero length” cross-linker, EDC, to distinguish among three feasible docking orientations of the spectrin dimer initiation site (Fig. 4B). As summarized above, the identified cross-links are incompatible with the docking orientations of the A and B models (Fig. 5). In contrast, the Docking C model is highly consistent with both the cross-links identified in the current study (Fig. 5C) and prior protease protection experiments (15). That is, cleavage sites unaffected by dimerization compared with monomer in exposed sites in our Docking C model (Fig. 6C, top panel), whereas sites protected in the heterodimer map to buried regions in the dimer interface (Fig. 6C, middle panel). Interestingly, several cleavage sites were shown to increase in yield in dimer compared with monomers; specifically the $\alpha_H$T2 and $\alpha_H$T3 sites become more sensitive to trypsin digestion in the dimer complex (15), presumably due to conformational changes induced by association of $\alpha_H$20–21 with $\alpha_H$1–2. These sites do map to relatively exposed regions in our heterodimer model, and therefore, substantial proteolytic cleavage of these sites in the dimer is consistent with model C (Fig. 6C, bottom panel).

As noted above, $\alpha$-actinin, a member of the spectrin superfamily, crystallized as an anti-parallel homodimer of two monomers, each consisting of four spectrin-like repeats (34). The four red cell spectrin repeats that are most homologous to $\alpha$-actinin are the repeats that constitute the spectrin dimer initiation site. Based on sequence homology, the $\alpha$-actinin dimerization site resembles two mirror image spectrin initiation sites. Hence it was reasonable to speculate that this analogy might extend to structural and binding properties of the dimerization sites of these two proteins; a hypothesis further supported by the fact that the 2 + 2 repeat interaction of the spectrin dimer initiation site has a $K_d$ in the low nanomolar range, whereas the
4 + 4 repeat α-actinin interaction has a $K_d$ in the low picomolar range (36–38). However, a prior study showed recombinant complementary α-actinin peptides could not dimerize with high affinity unless all four spectrin-type repeats were present (38). These relative binding affinities are roughly reflected in the amount of solvent accessible surface area buried by each of these complexes. That is, the interface between the two 4-repeat α-actinin monomers buries a total of $\sim 6244 \text{Å}^2$ of solvent-accessible surface area per dimer, and therefore a 2 + 2 repeat α-actinin interaction, which is very weak, buries $\sim 3122 \text{Å}^2$, while the low nanomolar 2 + 2 repeat spectrin interaction buries $\sim 3770 \text{Å}^2$ (1929 Å² for $\alpha_20–21$ and 1841 Å² for $\beta_1–2$). The only other dimeric structure of spectrin-type repeats is the crystalllographic structure of chicken brain α-spectrin repeats 15–17, which forms an anti-parallel homodimer (39). The interface between the two monomers buries a total of $\sim 3080 \text{Å}^2$ of solvent-accessible surface area for this 3 + 3 repeat interaction or $\sim 513 \text{Å}^2$ per individual repeat compared with 943 Å² per repeat in the high affinity spectrin dimer initiation site. It is currently unclear whether the CBo15–17 homodimer has a much smaller buried surface area, because it is not a physiologically heterodimer or because these and other spectrin repeats outside the dimer initiation site laterally associate with weak affinity.

In a separate study we recently used isothermal titration calorimetry to characterize binding interactions of α- and β-spectrin dimer initiation site peptides that contained the full dimer initiation site plus differing numbers of unpaired and laterally paired repeats ($\alpha_20–21/\beta_1–2$, $\alpha_20–21/\beta_1–3$, $\alpha_20–21/\beta_1–4$, $\alpha_18–21/\beta_1–2$, $\alpha_18–21/\beta_1–3$, and $\alpha_18–21/\beta_1–4$) (40). Using global analysis of these data, we determined the intrinsic enthalpy change and intrinsic entropy change involved in lateral pairing of the dimer initiation site as well as lateral pairing of up to two additional repeats next to the dimer initiation site. Heterodimerization of all combinations tested are both enthalpically and entropically driven. However, interaction of $\alpha_20–21$ and $\beta_1–2$ is dominated by a strongly favorable enthalpy change, at 30 °C, with a smaller contribution coming from a favorable entropy term. A large negative enthalpy change often implicates hydrogen bonding as a substantial contributor to the strength of the interaction. Because a favorable enthalpy change is the major component of the Gibbs free energy for the association of $\alpha_20–21$ and $\beta_1–2$, this suggests the inter-chain interaction force primarily comes from hydrophilic interactions. Our refined Docking C model reveals 10 H-bonds, 10 salt bridges, and a minor hydrophobic effect. It is often assumed that oppositely charged atoms in close proximity can form a salt bridge if they are less than or equal to 4.0 Å apart (41). In our model, we used a 6.0-Å cut-off for estimating intermolecular salt bridges due to likely lower accuracy of the model compared with crystallographic structures. In agreement with calorimetric results, the current model of the spectrin heterodimer initiation site indicates H-bonds and salt bridges are key forces in stabilizing spectrin dimer initiation.

In summary, nine intermolecular ion pairs in the $\alpha_20–21/\beta_1–2$ complex were identified using chemical cross-linking and characterization of resulting tryptic peptides by FT-ICR MS. These intermolecular distance constraints provide unambiguous experimental evidence for the docking interface within the spectrin dimer initiation site. Our model of the complex, which used these cross-link distance constraints to refine the most plausible homology model, provides a reliable medium resolution conformational model for this critical heterodimer complex. Structural features of this model are consistent with prior protease protection analyses, thermodynamic properties of heterodimer assembly for wild type subunits, and inability of α subunits containing the $\alpha_{145-147}$-6-residue deletion to heterodimerize. This structure is consistent with a mechanistic model for spectrin heterodimer assembly where long range complementary electrostatic interactions on the $\alpha_20–21/\beta_1–2$ repeats ensure alignment of the correct specific dimer initiation site repeats despite the fact both subunits have many tandem homologous repeats capable of improper low affinity interaction. After the long range electrostatic interactions within the dimer initiation site align these complementary repeats in proper register, this interaction is stabilized by extensive hydrophilic interactions in an interface that buries a substantial amount of surface area.

Acknowledgments—We thank Sandy Harper and Dr. John Rux for helpful discussions and critical reading of the manuscript, and Matthew Hart for performing the MALDI MS analyses. We also thank Dr. Cheng Luo for his assistance with the GRASP program.

REFERENCES

1. Nash, G. B., and Gratzer, W. B. (1993) *Biochemistry* **30**, 397–407
2. Mohandas, N., and Chasis, J. A. (1993) *Semin. Hematol.* **30**, 171–192
3. Bennett, V., and Baines, A. J. (2001) *Phys. Rev.** **B** 41, 1353–1392
4. Winkelmann, J. C., Chang, J. G., Tse, W. T., Scarpa, A. L., Marchesi, V. T., and Forget, B. G. (1990) *J. Biol. Chem.* **265**, 11827–11832
5. Sahr, K. E., Laurila, P., Kotula, L., Scarpa, A. L., Coupal, E., Leto, T. L., Linnenbach, A. J., Winkelmann, J. C., Speicher, D. W., Marchesi, V. T., Curtis, P. J., and Forget, B. G. (1990) *J. Biol. Chem.* **265**, 4434–4443
6. Mohandas, N., and Ax, N. S. (2006) *Transfus. Clin. Biol.* **13**, 29–30
7. Tse, W. T., and Lux, S. E. (2000) *Br. J. Haematol.* **104**, 2–13
8. Delaunay, I. (2007) *Blood Rev.* **21**, 1–20
9. Gallagher, P. G. (2004) *Semin. Hematol.* **41**, 142–164
10. Willmore, R., Marechal, J., Morle, L., Baklouti, F., Philippe, N., Kastally, R., Kotula, L., Delaunay, J., and Alloiso, N. (1993) *J. Clin. Invest.* **91**, 2091–2096
11. Speicher, D. W., and Marchesi, V. T. (1984) *Nature** **311**, 177–180
12. Speicher, D. W., Weglarz, L., and DeSilva, T. M. (1992) *J. Biol. Chem.* **267**, 14775–14782
13. Ursitti, J. A., Kotula, L., DeSilva, T. M., Curtis, P. J., and Speicher, D. W. (1996) *J. Biol. Chem.* **271**, 6636–6644
14. Speicher, D. W., DeSilva, T. M., Speicher, K. D., Ursitti, J. A., Hembach, P., and Weglarz, L. (1993) *J. Biol. Chem.* **268**, 4227–4235
15. Begg, G. E., Harper, L. S., Morris, M. B., and Speicher, D. W. (2000) *J. Biol. Chem.* **275**, 3279–3287
16. Young, M. M., Tang, N., Hempel, I. D., Gibson, B. W., and Dollinger, G. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5802–5806
17. Sinz, A. (2003) *J. Mass Spectrom.* **38**, 1225–1237
18. Yan, Y., Winograd, E., Viol, A., Cronin, T., Harrison, S. C., and Branton, D. (1993) *Science** **262**, 2027–2030
19. Speicher, K. D., Kolbas, O., Harper, S. L., and Speicher, D. W. (2000) *J. Biol. Chem.* **275**, 3279–3287
20. Arnold, K., Bordoli, L., and Schwede, T. (2006) *Bioinformatics* **22**, 195–201
21. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779–815
22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
Spectrin Heterodimer Initiation Site Model

23. DeLano, W. L. (2002) The PyMOL User’s Manual, DeLano Scientific, Palo Alto, CA
24. Jones, S., and Thornton, J. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 13–20
25. Guex, N., and Peitsch, M. C. (1997) Electrophoresis 18, 2714–2723
26. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) J. Am. Soc. Mass Spectrom. 14, 834–850
27. Comeau, S. R., Gatchell, D. W., Vajda, S., and Camacho, C. J. (2004) Bioinformatics 20, 45–50
28. Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyi, V., Mitchell, J. C., Nelson, E., Tsigeny, I., and Ten Eyck, L. F. (2001) Protein Eng. 14, 105–113
29. Chen, R., Li, L., and Weng, Z. (2003) Proteins 52, 80–87
30. Chen, R., and Weng, Z. (2003) Proteins 51, 397–408
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
32. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Protein Eng. 8, 127–134
33. Wilmotte, R., Harper, S. L., Ursitti, J. A., Marechal, J., Delaunay, J., and Speicher, D. W. (1997) Blood 90, 4188–4196
34. Ylanne, J., Scheffzek, K., Young, P., and Saraste, M. (2001) Structure 9, 597–604
35. Kalkhof, S., Ihling, C., Mechtler, K., and Sinz, A. (2005) Anal. Chem. 77, 495–503
36. Harper, S. L., Begg, G. E., and Speicher, D. W. (2001) Biochemistry 40, 9935–9943
37. Flood, G., Rowe, A. J., Critchley, D. R., and Gratzer, W. B. (1997) Eur. Biophys. J. 25, 431–435
38. Flood, G., Kahana, E., Gilmore, A. P., Rowe, A. J., Gratzer, W. B., and Critchley, D. R. (1995) J. Mol. Biol. 252, 227–234
39. Kusunoki, H., Minasov, G., Macdonald, R. I., and Mondragon, A. (2004) J. Mol. Biol. 344, 495–511
40. Li, D., Tang, H.-Y., and Speicher, D. W. (2007) Biochemistry 46, 10585–10594
41. Barlow, D. J., and Thornton, J. M. (1983) J. Mol. Biol. 168, 867–885
42. Honig, B., and Nicholls, A. (1995) Science 268, 1144–1149
43. El-Shafey, A., Tolic, N., Young, M. M., Sale, K., Smith, R. D., and Kery, V. (2006) Protein Sci. 15, 429–440