This is a postprint of an article published in
Wollert, T., Heinz, D.W., Schubert, W.-D.
Thermodynamically reengineering the listerial invasion complex InlA/E-cadherin
Thermodynamically Re-engineering the Listerial Invasion Complex InlA / E-Cadherin

Thomas Wollert¹, Dirk W. Heinz², Wolf-Dieter Schubert¹,³

¹Molecular Host-Pathogen Interactions, Division of Structural Biology
²Division of Structural Biology, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig, Germany.
³To whom correspondence should be addressed. Wolf-Dieter Schubert, Molecular Host-Pathogen Interactions, Division of Structural Biology, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig, Germany E-mail: Wolf-Dieter.Schubert@helmholtz-hzi.de, Phone: +49 531 6181 7043, Fax: +49 531 6181 7099.

Manuscript information: 17 pages text, 3 Figures, 207 words in abstract, 37724 characters text

Keywords: Structure-based re-engineering, thermodynamics, Listeria monocytogenes, Internalin A, E-cadherin
Abstract

Biological processes essentially all depend on the specific recognition between macromolecules and their interaction partners. Although many such interactions have been characterized both structurally and biophysically, the thermodynamic effects of small atomic changes remain poorly understood. Based on the crystal structure of the bacterial invasion protein internalin (InlA) of *Listeria monocytogenes* in complex with its human receptor E-cadherin (hEC1), we analyzed the interface to identify single amino acid substitutions in InlA that would potentially improve the overall quality of interaction and hence increase the weak binding affinity of the complex. Dissociation constants of InlA-variant/hEC1 complexes, as well as enthalpy and entropy of binding were quantified by isothermal titration calorimetry. All single substitutions indeed significantly increase binding affinity. Structural changes were verified crystallographically at 2.0 Å resolution or better, allowing thermodynamic characteristics of single substitutions to be rationalized structurally and providing unique insights into atomic contributions to binding enthalpy and entropy. Structural and thermodynamic data of all combinations of individual substitutions result in a thermodynamic network allowing the source of cooperativity between distant recognition sites to be identified. One such pair of single substitutions improves affinity 5000-fold. We thus demonstrate that rational re-engineering of protein complexes is possible by making use of physically distant hot spots of recognition.
The Gram-positive bacterium *Listeria monocytogenes* causes severe infections in immuno-compromized individuals and unborn fetuses (1). As part of its invasion strategy, *L. monocytogenes* is able to breach the intestinal barrier by inducing its own uptake into normally nonphagocytic cells using the invasion protein internalin (InlA) (2). Structurally InlA consists of an N-terminal cap, a LRR and an immunoglobulin-like interrepeat (IR) domain (3, 4), followed by three spacer domains to allow presentation on the cell surface (5). E-cadherin, the most abundant protein in epithelial-cell adherens junctions, is crucial in embryogenesis (6) and in maintaining epithelial integrity (7). It consists of five extracellular, immunoglobulin-like domains (EC1-5), a transmembrane α-helix and an intracellular domain linked to the actin cytoskeleton (8). The N-terminal domain hEC1 is responsible for cell-cell contacts (9) and is also the receptor of InlA (10, 11). The crystal structure of the functional domain of InlA in complex with hEC1 revealed that InlA binds hEC1 through the concave face of its LRR-domain (5). Despite burying 2400 Å² of solvent-accessible surface area upon complex formation, the binding affinity ($K_D = 8 \pm 4 \mu M$) is rather weak. However, weak affinity does not correlate with low binding specificity, as indicated by a narrow range of EC1-domains of other species recognized by InlA (11).

Manipulating protein-protein interaction surfaces to increase binding affinity or to change binding specificity still represents a major challenge. Essentially two competing approaches exist, based either on selection of efficient binders from large, randomly created libraries (12) or on computational design (13). The latter, however, requires comprehensive knowledge of protein recognition and underlying physical mechanisms, which have not yet been fully analyzed or quantified (14). Unresolved questions include (i) enthalpic and entropic contributions to binding affinity, (ii) the source of enthalpy-entropy compensation phenomena (15), (iii) the precise role of water molecules within protein-protein interaction surfaces (16) and (iv) cooperativity between adjacent and distant interaction sites (17).

To contribute to a better understanding of protein-protein interactions we have applied a rational protein design approach without relying on computational methods. Instead, the high resolution crystal structure of the InlA/hEC1 complex allowed single amino acid substitutions in InlA to be identified that potentially increase its binding affinity for hEC1. We characterized the binding affinity of InlA-variants carrying single or double substitutions by isothermal titration calorimetry (ITC) deriving precise enthalpic and entropic contributions to complex formation. We have similarly solved the crystal structures of InlA-variants in complex with hEC1, yielding precise structural data on changes introduced through the substitutions. Thermodynamic properties may thus be directly correlated with structural
changes, providing unique insights into enthalpic and entropic contribution of single amino acid side chains to macromolecular complex formation and to the cooperative behavior of combinations of single mutants.

Results

Structure-based mutant design

Though the crystal structure of the InlA/hEC1 complex (5) (Fig. 1A) indicates a comparatively large (18) interaction interface (2400 Å²), the binding affinity of InlA for hEC1 at KD = 8 ± 4 µM is surprisingly weak, presumably due to low surface complementarity and a large number of bridging water molecules. Using the structural information, we designed the following InlA-variants to increase the binding affinity of the complex:

- **Tyr369Ala** (Y369A): In uncomplexed InlA, Tyr369 forms a well-ordered stacking interaction with Tyr347 (Fig. 2A). In the complex InlA/hEC1, Asn27hEC1 displaces Tyr369 causing it to swing around its χ₁-angle away from its stacking interaction with Tyr347, and displacing Asn370 and His392 from a similar stacking interaction with Phe348. Replacing Tyr369 by alanine would eliminate this rearrangement of surface residues.

- **Tyr369Ser** (Y369S): While an alanine at position 369 eliminates the unfavorable conformation of Tyr369 (above), its small size and lack of hydrogen bond donor or acceptor groups would prevent it from directly interacting with hEC1. Placing serine at this position should allow direct or water-mediated hydrogen bonds to hEC1.

- **Ser192Asn** (S192N): Ser192InlA forms a water-mediated hydrogen bond to hEC1. As a direct hydrogen bond between InlA and hEC1 would increase binding affinity, we replaced Ser192 by slightly longer asparagine allowing it to potentially bridge the gap to hEC1.

- **Gly194Ser+Ser** (G194S+S): Compared to other LRRs of InlA, repeat 6 consists of 21 residues instead of the canonical 22. This shortens the loop after the LRR β-strand (5), discontinues the asparagine ladder characteristic of LRR-proteins (19), and creates a 7.5 Å-wide, hydrophobic, water-filled cavity on the surface of InlA. To restore the regular LRR architecture, an additional serine (+S) was introduced after Gly194, while Gly194 itself was replaced by serine, the most common residue at this position in other LRRs of InlA.

Structural verification of predicted atomic-scale changes

InlA-mutant/hEC1 complexes were analyzed by X-ray crystallography at 2.0 Å resolution or better. For data collection and refinement statistics see Table S1. Superimposing all
complexes indicates that single substitutions in InlA do not affect the structure of InlA itself, nor the geometry of the complex (r.m.s.-deviations 0.65 Å). This allows atomic changes in the immediate vicinity of the mutation to be analyzed, especially as regards hydrogen-bond networks and water-mediated interactions.

**Y369A and Y369S:** As postulated, the substitution of Tyr369 with alanine or serine allows Asn370 and His392 to maintain their stacking interaction with Phe348 as in uncomplexed InlA (Fig. 2B). Water molecules near Tyr369 in InlA/hEC1 are largely conserved in Y369A/hEC1 and Y369S/hEC1 (black spheres). One such water molecule, hydrogen bonded to Asn27-Nδ2 in all complexes, additionally forms a second hydrogen bond to Ser369-Oγ in Y369S/hEC1, bridging InlA-Y369S and hEC1. Two further water molecules (red) bound by Asn370 and His392 are present only in Y369A/hEC1 and Y369S/hEC1. They replace a water molecule from the second solvation shell of InlA/hEC1 (blue).

**S192N:** In InlAwt/hEC1, Ser192 adopts two distinct, equally occupied conformations, each involved in a water-mediated hydrogen-bond to the main-chain oxygens of Phe17hEC1 or Pro18hEC1 (blue residues, Fig. 2C). The first of these water molecules additionally interacts with Ser172InlA, the second with Asp213InlA. Engineered Asn192InlA displaces the first bridging water, introducing a direct hydrogen-bond from Asn192InlA-Nδ2 to Phe17hEC1-O. A low B-factor of the second, now Asn192InlA-coordinated water (dark red sphere in Fig. 2C) indicates interaction of Asn192InlA, Pro18hEC1 and Asp213InlA in S192N/hEC1 to be much tighter than in InlA/hEC1. Asn192InlA-Oδ1 intramolecularly hydrogen-bonds to backbone atoms of neighboring repeats preventing additional stabilizing contacts to hEC1.

**G194S+S:** Replacing Gly194InlA by serine and inserting a second serine after Ser194 (+S), allows LRR6 to adopt a structure similar to that of all other repeats (Fig. 2D). Correspondingly, Asn195 flips into the hydrophobic core complementing the asparagine ladder. In addition, the hydrophobic water-filled cavity between wild-type InlA and hEC1 (Fig. 2E) is eliminated, reducing the distance between the two proteins from ~10 to ~4 Å (Fig. 2F). At least four water molecules are displaced from the interface (light-blue spheres in Fig. 2D). One water molecule, hydrogen bonded by residue +S additionally forms a long range interaction (4.0 Å) to Glu54hEC1, while a second water, hydrogen bonded to both Glu54 and Lys61hEC1 similarly forms a long range interaction of 4.2 Å to +S (Fig. 2D).

**Complex formation of InlA/hEC1 is enthalpy and entropy driven**

Using isothermal titration calorimetry (ITC), we have narrowed down the dissociation constant for wild-type InlA/hEC1, found to be $8 \pm 4$ µM by analytical ultracentrifugation (5),
to $3 \pm 1 \mu M$. In addition, ITC allows changes in binding affinity to be separated into enthalpic and entropic contributions.

Analyzing the association of InlA<sub>wt</sub> and hEC1 in different buffers, indicates that the apparent enthalpy ($\Delta H_{\text{app}}$) of complex formation is dependent on the ionization enthalpy ($\Delta H_{\text{ion}}$) of the buffer (20), implying that complex formation is associated with an exchange of protons. $\Delta H_{\text{app}}$ is found to be $-6.7 \pm 0.3 \text{ kJ/mol}$ in cacodylate ($\Delta H_{\text{ion}} = -2.5 \text{ kJ/mol}$), $-1.8 \pm 0.2 \text{ kJ/mol}$ in Heps ($\Delta H_{\text{ion}} = 23.9 \text{ kJ/mol}$), and $3.5 \pm 0.2 \text{ kJ/mol}$ in Tris ($\Delta H_{\text{ion}} = 47.7 \text{ kJ/mol}$). Plotting $\Delta H_{\text{app}}$ against $\Delta H_{\text{ion}}$ (not shown) indicates that $0.2 \pm 0.1$ protons (gradient) are taken up during complex formation. Though the side-chain involved remains unclear, the binding enthalpy may be corrected for $\Delta H_{\text{ion}}$. $\Delta H_{\text{ion}}$-independent binding enthalpy is thus $-6 \text{ kJ/mol}$ ($\Delta H_{\text{ion}}=0$) and entropy ($T\Delta S$) is $25 \text{ kJ/mol}$. Complex formation of InlA and hEC1 is thus both entropy and enthalpy driven.

### Revealing atomic contributions to binding enthalpy and entropy

To compare enthalpic and entropic contributions to complex formation of InlA-variants and InlA<sub>wt</sub>, ITC experiments were performed in Heps buffer and 20mM CaCl<sub>2</sub>. Values were corrected for $\Delta H_{\text{ion}}$ to place them on an absolute scale (Fig. 3). Differences in thermodynamic quantities remain unaffected as they are independent of $\Delta H_{\text{ion}}$. Surprisingly, the atomic modifications to InlA result in unexpectedly large and divergent changes in the thermodynamics of complex formation:

**Y369A and Y369S:** Both substitutions contribute enthalpically to complex formation: $\Delta \Delta H = -13 \text{ kJ/mol}$ for Y369A (green boxes in Fig. 3) and $-11 \text{ kJ/mol}$ for Y369S (turquoise in Fig. 3, Fig. S1). In Y369A the favorable enthalpic contribution is counteracted by an unfavorable reduction in binding entropy ($\Delta T\Delta S = -8 \text{ kJ/mol}$), a case of “enthalpy/entropy compensation” (15). Compared to InlA<sub>wt</sub>, Y369A therefore increases binding affinity ($K_D = 400 \pm 100 \text{ nM}$) to hEC1 only 7.5-fold. In Y369S (blue labels in Fig. 3) a much smaller entropic compensation ($\Delta T\Delta S = -2 \text{ kJ/mol}$) results in the highest binding affinity for hEC1 ($K_D = 90 \pm 20 \text{ nM}$) of any single InlA-mutant investigated in this study.

**S192N and G194S+S** (magenta and yellow boxes in Fig. 3): Whereas Y369A and Y369S favorably decrease the enthalpy of complex formation compared to wild type InlA, both S192N and G194S+S unfavorably increase this contribution ($\Delta \Delta H = 5$ and $11 \text{ kJ/mol}$ for S182N and G194S+S). In the case of G194S+S the increase in enthalpy is sufficient to make complex formation endothermic ($\Delta H = 5 \text{ kJ/mol}$). The increase in binding enthalpy of both
substitutions is, however, more than compensated by a large favorable increase in binding entropy ($\Delta T \Delta S_{S192N} = 12$ kJ/mol, $\Delta T \Delta S_{G194S+S} = 19$ kJ/mol) resulting in a significantly higher binding affinity for hEC1 ($K_D = 0.20 \pm 0.05 \mu$M) than InlA$^{wt}$. Both S192N and G194S+S thus improve surface complementarity of InlA/hEC1 allowing the entropically favorable elimination of around one and four constrained water molecules from the interface, respectively.

**Thermodynamics of long-range cooperativity between combined mutations**

To achieve higher binding affinity, the described amino acid substitutions were combined to yield the four InlA-variants S192N-Y369A, S192N-Y369S, G194S+S-Y369S and S192N-G194S+S (bottom row, Fig.3). Our data indicate that the combination of physically distant, single substitutions significantly increases the binding affinity for hEC1. Thus the binding affinities of S192N-Y369A, S192N-Y369S, and G194S+S-Y369S are 200, 2500, and 5000-fold (orange boxes, Fig. 3) that of InlA$^{wt}$. Though the sites of substitution are separated by ~34 Å (Fig. 1A), binding affinities of individual substitutions are not merely additive but indicate positive cooperativity instead. ‘Synergy factors’ were calculated by dividing the increase in binding affinity of the doubly substituted variant (lower colored box in Fig. 3) by that of the single substitution variant (upper box of identical color). For S192N-Y369A this amounts to $13/7.5 \approx 27/15 \approx 1.8$, while S192N-Y369S and G194S+S-Y369S both yield a value of ~5 (Fig. 3). The synergy, though precise in terms of binding affinities, is less well defined with respect to enthalpy or entropy – in common with previous observations (14). Thus, $\Delta \Delta H$ and $\Delta T \Delta S$ for identical substitutions introduced either into InlA$^{wt}$ or combined with an InlA-variant, differ by no more than 3 kJ/mol, only slightly larger than the average experimental error of 1-2 kJ/mol. The synergy, however, appears to be linked to an increase in entropy, as $\Delta T \Delta S$ is always 2-3 kJ/mol higher for double substitution variants than for corresponding single substitution variants (boxes of identical color in lower and upper rows of Fig. 3) – in contrast to differences of 0 to -1 kJ/mol for $\Delta \Delta H$.

The tightest binding affinity of an InlA-variant for hEC1 is that of G194S+S-Y369S, $K_D = 0.6 \pm 0.2$ nM. Only two rationally chosen substitutions in InlA thus suffice to transform the weak binding affinity of the wild-type complex, $K_D = 3 \pm 1$ µM, to a tight fit comparable to that of typical proteinase/proteinaceous-inhibitor complexes (21) - one of the highest increases in binding affinity (5000-fold) reported for any protein-protein interaction (22).

In contrast to the synergy for S192N-Y369A, S192N-Y369S, and G194S+S-Y369S, the substitutions of the fourth double variant S192N-G194S+S are anti-cooperative. The synergy
factor is $0.8/15 \approx 1.6/30 = 0.05$ (Fig. 3) resulting in a binding affinity that is similar to that of the individual substitution variants rather than ~10-fold stronger as expected if the effects were additive. An increase of $\Delta T\Delta S$ of $\sim 7$ kJ/mol indicates this to be an entropic effect.

**Discussion**

5 **Thermodynamics of complex formation**

Despite complex formation of InlA and hEC1 being both enthalpically and entropically favored, and an apparent large interaction surface, the binding affinity of the complex is weak. Thermodynamically, binding entropy (dominated by exclusion of water molecules) outweighs binding enthalpy (hydrophilic interactions). Only two hydrophobic contact areas, centered on Val$_{3\text{hEC1}}$ and Pro$_{16\text{hEC1}}$, exist in InlA/hEC1. Nevertheless, by excluding numerous rotationally-restrained water molecules during complex formation (5), these hydrophobic interactions would appear to contribute substantially to the favorable increase in binding entropy of $\Delta T\Delta S = 25 \pm 1$ kJ/mol. The small enthalpic contribution to complex formation ($\Delta H = -6 \pm 0.2$ kJ/mol) in turn correlates structurally with the paucity of enthalpically favorable direct contacts between the proteins. These include only seven hydrogen bonds, three salt bridges, and eight water bridged interactions (5). Compared to tighter protein complexes, InlA/hEC1 retains significantly more water molecules within the interface (16).

On the other hand, the low binding affinity and poor surface complementarity of InlA/hEC1 provides us with an optimal system to study the crucial role of water in complex formation. By excluding more water molecules during complex formation, we may entropically stabilize the interaction. Alternatively, water molecules enthalpically contribute to binding affinity if their hydrogen bonding potential is optimized to bridge hydrophilic interfaces (16). The high-resolution structural data (Fig. 1-2) for InlA-variant/hEC1 complexes may thus be used to interpret the observed changes in thermodynamic parameters (Fig. 3).

Note, however, that there are limitations to this approach, in particular as it is clearly impossible to take all contributing factors (including distant water molecules) into account. The tendency of enthalpy or entropy to compensate a change in the other (enthalpy-entropy compensation) especially in weak intermolecular interactions (23), furthermore, affects experimentally determined changes in enthalpy and especially entropy. As a result, small changes, particularly in entropy, are difficult or impossible to interpret in a structural sense. Large changes in thermodynamic contributions are not affected to the same extent allowing their cause to be discussed qualitatively in terms of structural change even though their
constituent contributions are not precisely resolved. Examples include the rotational entropies of amino acid side chains, shown to be amenable to calculation (24), and the established entropic contribution of excluding water molecules from interfaces (23).

**Y369A and Y369S** (green and turquoise labels in Fig. 3): Compared to InlA/hEC1, the substitutions Y369A and Y369S improve the enthalpy of binding by $\Delta\Delta H = -13$ or $-11$ kJ/mol. Structurally, this may be rationalized by the enthalpically favorable stacking (25) of Phe348, Asn370 and His392 being retained in the variant complexes from uncomplexed InlA (transparent, pink residues in Fig. 2A) rather than being disrupted as in InlA/hEC1 (blue in Fig. 2A). This example confirms the general view that pre-organized interaction surfaces in the unbound state play a major role in protein-recognition (26).

The favorable increase in enthalpy of Y369A and Y369S in complex formation (above) is offset by a reduction in the entropic contribution by $\Delta T\Delta S = -8$ or $-2$ kJ/mol compared to that of wild-type InlA. These changes in entropy may substantially be affected by solvent entropy compensating the change in enthalpy, thereby limiting their detailed correlation with structural data (27). In a qualitative sense, the loss in entropy in Y369A/hEC1 ($\Delta T\Delta S = -8$ kJ/mol) and Y369S/hEC1 (-2 kJ/mol) appears to correlate with the size of a hydrophobic patch, being exposed as a result of the substitutions.

**S192N**: Ser192 adopts two distinct conformations in InlA wt/hEC1 (Fig. 2C and Fig. S2A) each of which hydrogen bonds a bridging water molecule (see above). Replacing Ser192 by asparagine displaces one of these water molecules. Excluding a single water molecule from a protein interface leads to a favorable increase in solvation entropy of $\sim 6$-9 kJ/mol (23). The observed exclusion of a water molecule thus presumably contributes to the 12 kJ/mol increase in binding entropy of S192N/hEC1 (magenta boxes in Fig. 3). The loss of the water-mediated hydrogen-bond Ser172 InlA–Phe17hEC1 in turn may explain the corresponding, unfavorable increase in binding enthalpy ($\Delta\Delta H = 5$ kJ/mol).

**G194S+S** fills a large depression on the surface of InlA (compare Fig. 2E and 2F) excluding four rotationally restrained water molecules from the interface and dramatically increasing binding entropy ($\Delta T\Delta S = 19$ kJ/mol). Why does the binding enthalpy increase to such an extent as to make complex formation endothermic? In the InlA wt/hEC1, a distance of 10Å between InlA and hEC1 is sufficiently large to allow bulk solvent to fill the cavity between the two independently solvated surfaces (Fig. 2E). In G194S+S/hEC1, the distance is reduced to $\sim 4$ Å. This distance appears too narrow to allow independent solvation of each surface yet too wide for a single bridging layer of solvent. Instead the inter-protein hydrogen-bonding
network is observed to be discontinuous in the crystal structure which would enthalpically be unfavorable.

**Synergy of combined substitutions**

Strikingly, our study indicates that changes in binding affinity of single substitutions are not simply additive when the substitutions are combined in a single protein.

For the combination of S192N and G194S+S, we observe anti-cooperative behavior characterized by a synergy factor of 0.05. Binding affinity of InlA$^{S192N-G194S+S/hEC1}$ is thus weaker than the combination of individual substitutions would imply. The effect appears largely entropic, as $\Delta T\Delta S_{S192N-G194S+S} \approx -7$ kJ/mol (Fig. 3), implying that overall about one water molecule less is displaced by S192N-G194S+S than by S192N and G194S+S combined. The crystal structure of S192N-G194S+S/hEC1 correspondingly indicates that the side chain of Asn192InlA is locked into a tight intramolecular hydrogen bond to the physically adjacent backbone nitrogen of Ser194 (Fig. S2B-C), preventing Asn192 from displacing a water molecule as described for S192N/hEC1 (compare Fig. S2B-C with D).

Note that this explanation, as well as those below, only reflect structural changes obvious in the corresponding crystal structures. More subtle contributions to the dynamics of protein association or to the structure and stability of the unbound proteins brought about by substituting individual residues fall outside the scope of this publication.

Potentially more interesting than the negatively cooperative S192N-G194S+S are the double variants S192N-Y369A, S192N-Y369S and G194S+S-Y369S, all of which are characterized by synergy factors above one, indicating individual substitutions to be cooperative. As indicated above, these observed synergies are unambiguous in terms of binding affinities. The enthalpic (1–2 kJ/mol) and entropic (2–3 kJ/mol) contributions are, however, similar or only slightly larger than the experimental error (1–2 kJ/mol), and should hence not be over-interpreted. Furthermore, it is interesting to note that S192N-Y369A, S192N-Y369S and G194S+S-Y369S all combine individual substitutions physically separated by more than 30 Å (Fig. 1A). Previous studies of protein-complexes had indicated synergy to be limited to clustered residues, whereas spatially distant improvements were strictly additive (28). Only recently has this been challenged by the report of positive cooperativity for substitutions in the T-cell receptor variable domain separated by 20 Å (17).

Structurally, the observed long-range synergy in InlA-variants may be due to a physical link in the form of a β-strand between two sites of substitution. The variants S192N and Y369A/S increase binding affinity of hEC1 through favorable interactions to Phe17$hEC1$ and Asn27$hEC1$.
respectively, located at either end of β-strand b (βb, residues 19-26) of hEC1 (Fig. 1). In InlA/hEC1 β-strand βa (residues 2-10) and loop βa-βb (residues 10-19) constitute the major part of the interface, whereas interactions of βb to InlA are restricted to two water-mediated contacts. By stabilizing either end of βb, substitutions S192N and Y369A/S appear to stabilize hEC1 and hence the interface as a whole, resulting in the observed synergistic increase in binding affinities (Fig. 1 and 3).

The reason for a synergy factor of 1.8 for S192N-Y369A, compared to 5 for S192N-Y369S, may structurally be linked to the presence of an additional hydrogen bond to the C-terminal end of βb in Y369S (structure of S192N-Y369S/hEC1). This direct interaction is absent in Y369A potentially indicating a lower stabilization of βb of hEC1 and reducing synergy to 1.8.

In the case of G194S+S-Y369S, the insertion of serine (+S) restores the canonical LRR-architecture (Fig. 2D and Fig. S2C) and fills a large cavity between LRR6 of InlA and residues Glu54hEC1 and Lys61hEC1. As the surface of G194S approaches the corresponding hEC1 surface more closely at least four intervening water molecules are physically excluded. Long-range interactions to Glu54hEC1 and Lys61hEC1 appear to stabilize the complex. Glu54 and Lys61 are respectively located in β-strands βd and βe of hEC1 (Fig. 1B), that form a β-sheet with strand βb. The stabilization of β-strands βd and βe appears to be transmitted through a β-sheet to βb and thereby give rise to a measurable positive cooperativity. This mechanism of long-range cooperativity may be equivalent to the intra-molecular, allosteric effects observed in multimeric protein complexes (29).

**Methods**

**Mutagenesis, Expression and Purification**

For structural and biophysical studies, functionally relevant fragments of InlA (residues 36-496) and the first extracellular domain of E-cadherin (hEC1, residues 1-105) were used (5). Site-directed mutations were introduced using QuikChange mutagenesis (Stratagene). Protein expression and purification were as described (5).

**Isothermal titration calorimetry**

Isothermal titration calorimetry (ITC) was performed using a MCS calorimeter (MicroCal Inc.). All samples were dialyzed against 50mM HEPES (pH 7.5) and 20mM CaCl₂. Concentrations were determined spectrophotometrically at 280nm. Titrations were performed at 20°C by injecting 5-10µl aliquots of wild-type or mutant InlA into the ITC cell containing 1.35ml of hEC1. Data were corrected for heat of dilution (ΔHdil). Binding stoichiometry,
enthalpy, as well as equilibrium association constants were determined using the “single set of independent sites” model of molecular association (MicroCal Origin 2.9).

To determine the influence of the ionization enthalpy (ΔH$_{ion}$) of the buffer on the apparent enthalpy of binding (ΔH$_{app}$), complex formation of wild-type InlA and hEC1 was recorded in 50mM cacodylate and in 50mM Tris/HCl buffers supplemented with 20mM CaCl$_2$. The relationship between ΔH$_{ion}$ and ΔH$_{app}$ was analyzed by linear regression to evaluate the number of protons exchanged during complex formation. The binding enthalpy was corrected for the ionization enthalpy of the buffer (ΔH$_{ion}$ = 0 intercept of ΔH$_{app}$ plotted against ΔH$_{ion}$).

**Crystallization and data collection**

Complexes of InlA-variants and hEC1 were crystallized by hanging drop vapor diffusion. Total protein concentration was 5mg/ml, stoichiometry of InlA:hEC1=1:1. The reservoir solution contained 20-25% PEG4000, 100mM MES/Tris buffer (pH 7.0-7.5), 100mM Na acetate and 20-100mM CaCl$_2$. 20% PEG400 (v/v) was added to the reservoir solution for cryo-protection.

**Structure determination**

X-ray data were collected using synchrotron radiation and MARCCD detectors (Marresearch, Germany) at beamlines BW6 (Y369A/hEC1 and S192N/hEC1, λ = 1.05 Å) and X13, EMBL, DESY, Hamburg, Germany (G194S+S/hEC1, λ = 0.80 Å), and BL1, Protein Structure Factory, BESSY, Berlin, Germany (S192N-Y369S/hEC1, λ = 0.95 Å). Data for complexes G194S+S-Y369S/hEC1 and S192N-G194S+S were collected using a rotating anode (Rigaku, 1.54 Å) and an image plate detector (R-Axis IV). Data were processed using HKL (30), XDS (31) and the CCP4 suites (32). All structures were solved by molecular replacement using EPMR (33) and the wild-type InlA/hEC1 complex as search model (5). REFMAC5 (34) was used for refinement, Coot (35) for model building, structural analysis and structure validation, as was WHATIF (36). Figures were prepared using PYMOL (www.pymol.org).

**Acknowledgement**

We thank our colleagues Dr. Uwe Kärst and Prof. Jürgen Wehland (Division of Cell Biology) for the isothermal titration calorimeter. We express our appreciation to the staff of beamlines BW6 and X13 (EMBL, DESY, Hamburg, Germany), and beamline BL1 (Protein Structure Factory, BESSY, Berlin, Germany) for access to beam time and for assistance during X-ray data collection. Financial support from the Deutsche Forschungsgemeinschaft as part of the Priority Program SPP1150 (to W.D.S, SCHU 1560/1-1/1-2) is gratefully acknowledged.
Authors Contribution: T.W. and W.-D.S. designed research; T.W. performed research; T.W., D.W.H., and W.-D.S. analyzed data; and T.W. and W.-D.S. wrote the paper.

The authors declare that they have no conflicts.

Data Deposition: coordinates of structures have been deposited with the Protein Data Bank (accession codes 2OMT, 2OMU, 2OMX and 2OMZ).
References


Figure legends

**Figure 1:** Overall structure of InlA / hEC1 (dark / light blue). (A) Re-engineered residues are indicated by spheres. Though separated by 34 Å, combinations of substitutions from entropically (pink) and enthalpically (orange) dominated "hot spots" act synergistically by stabilizing β-strand b of hEC1 (red). (B) Close up view of the interaction interface. S192N and Y369A/S (ball-and-sticks) stabilize opposite ends of βb. G194S+S shortens the distance to residues Glu54hEC1 and Lys61hEC1 (ball-and-stick) in βd and βe, respectively. Stabilization is transmitted through β-sheet bde to the N-terminus of βb. All structural figures were prepared using Pymol (www.pymol.org).

**Figure 2:** Structural details of InlA mutations.

(A) Tyr369InlA-induced side-chain rearrangements during complex formation. Superposition of uncomplexed InlA (pink) and InlA/hEC1 (blue). On complex formation Ans27hEC1 (light blue, top) causes Tyr369 InlA (dark blue) to flip to an alternative, less favorable conformation, displacing Asn370 and His392 from their stacking interaction with Phe348 (black arrows).

(B) The interface near Tyr369InlA in InlA/hEC1 (blue), Y369A/hEC1 (green) and S192N-Y369S/hEC1 (orange). Molecular surface of Y369A is in grey, that of hEC1 in pink. Side-chains changing conformation during complex formation are shown for InlA/hEC1 (blue) and S192N-Y369S/hEC1 (dark orange). Spheres represent water molecules: black - conserved in all complexes, orange - present in mutant complexes, blue - present only in the wild-type complex. Y369A/S prevents the disruptive re-orientation of Tyr369 during complex formation exposing a water-filled cavity. Ser369-Oγ binds two conserved water molecules, one also coordinated by Asn27hEC1-Oδ1.

(C) Ser192InlA in InlA/hEC1 (blue) and S192N/hEC1 (dark red/pink). The two conformations of Ser192 respectively form water-bridged hydrogen-bonds to Phe17hEC1-O or Pro18hEC1-O. The water molecules are additionally hydrogen bonded by Asp213InlA and Ser172InlA. Substituting Ser192 by asparagine displaces one of the water molecules and introduces a direct hydrogen bond to Pro18-O. The second water molecule maintains the hydrogen bonding pattern of the wild-type complex.

(D-F) InlA/hEC1 (blue) and G194S+S/hEC1 (pink). (D) The interaction of InlA-LRR5 and 6 with hEC1. The mutation G194S and the insertion of an additional serine (+S) restores the
canonical LRR-repeat geometry in LRR6, flipping Asn195 (arrow) into the hydrophobic core of InlA (dark/light grey, wild-type/variant) and removing a large, water-filled cavity.

(E) Electron density (1σ-contoured 2Fo−Fc, green - protein, red - water) of LRR6 in InlA/hEC1 (5). The 21-residue LRR6 creates a hydrophobic cavity between Gly194$_{\text{InlA}}$ and Glu54$_{\text{hEC1}}$ and Lys61$_{\text{hEC1}}$. Water molecules filling the gap are poorly defined (weak electron density).

(F) The equivalent view as (E) for G194S+S/hEC1. The gap between interaction partners is narrower, yielding a well-defined yet unsaturated water cluster.

**Figure 3: Thermodynamic network.** Binding affinities (K$_D$), binding enthalpies (ΔH) and entropies (TΔS) are summarized for wild-type and variant complexes of InlA with hEC1 (T = experimental temperature, 293 K). Enthalpies have been corrected for ΔH$_{\text{ion}}$ of InlA$_{\text{wt}}$/hEC1 to place them on an absolute scale prior to calculating TΔS-values. Substitutions and associated changes in binding enthalpies (ΔΔH = ΔH$_{\text{var}}$ − ΔH$_{\text{wt}}$) and entropies (ΔTΔS = TΔS$_{\text{var}}$ − TΔS$_{\text{wt}}$) are listed in color-coded boxes: green - Y369A, magenta - S192N, turquoise - Y369S, yellow - G194S+S. Blue and red fonts denote thermodynamically favorable and unfavorable changes. Note the excellent reproducibility for identical substitutions introduced into different backgrounds. Combining mutations results in super-additive strengthening of corresponding complexes (orange-colored arrow). Synergy factors (bottom row, see text for definition) indicate positive (>1) and negative (<1) cooperativity between mutations. Ribbon pictograms mark complexes for which crystal structures have been solved. The color-graded arrow indicates binding affinities of InlA-variants (blue ~10 µM and red ~1 nM K$_D$).
InlA
hEC1
β
cooperativity
123
4
5
6
7
8
9
10
11
12
13
14
15
34 Å
S192N
G194S+S
Y369A/S
A

Wollert et al., 2007
Figure 2

Wollert et al., 2007
Figure 3

Wollert et al., 2007

Synergy factor

1.8

5

0.05

5
Figure S1

**A** Y369S

- **Time (min)**
  - 0, 50, 100, 150
- **µcal/sec**
  - -0.6, -0.3, 0.0
- **kcal/mole of injectant**
  - Exothermic
- **Molar Ratio**
  - 0, 1, 2

**B** G194S+S

- **Time (min)**
  - 0, 50, 100
- **µcal/sec**
  - 0, 0.2, 0.4
- **kcal/mole of injectant**
  - Endothermic
Figure S2

(A) InlA wild-type

(B) InlA S192N

(C) InlA G194S+S

(D) InlA S192N-G194S+S
Figure S1: Isothermal titration curves. (A) The titration curve of hEC1 against Y369S demonstrates the association to be exothermic. 8 µl aliquots of Y369S (0.43 mM) were successively injected into a 0.04 mM solution of hEC1. The association affinity constant and the binding enthalpy were derived by curve fitting using the single set of independent sites model. (B) Titrating hEC1 (0.035 mM) against G194S+S (0.66 mM) using 5 µl injections reveals an unusual endothermal association reaction. Both titrations yield a binding stoichiometry of 1.0 ± 0.1.

Figure S2: Schematic representation of local rearrangements in InlA-variants. Leucine-rich-repeat (LRR) 5-7 of InlA-variants are shown in green, interacting residues of hEC1 in yellow. Amino acid substitutions are highlighted in red.

(A) InlA/hEC1: disordered Ser192\textsubscript{InlA} and its water-mediated interactions to Phe17\textsubscript{hEC1}-O and Pro18\textsubscript{hEC1}-O. Two residues of neighboring repeats, Asp213\textsubscript{InlA} and Ser172\textsubscript{InlA}, additionally stabilize the water cluster near Ser192\textsubscript{InlA}. The shortened LRR6 excludes Asn195\textsubscript{InlA} from the canonical asparagine ladder within the hydrophobic core and creates a large depression facing Glu54\textsubscript{hEC1} and Lys61\textsubscript{hEC1}.

(B) Substituting Ser192\textsubscript{InlA} by asparagine creates a direct hydrogen bond to Phe17\textsubscript{hEC1} by excluding one water molecule previously stabilized by Ser172\textsubscript{InlA}.

(C) Restoring the canonical LRR-architecture by inserting a serine after position 194 and replacing Gly194\textsubscript{InlA} by serine flips Asn195 into the hydrophobic core to complete the asparagine ladder and reduces the distance to Glu54\textsubscript{hEC1} and Lys61\textsubscript{hEC1}.

(D) Combination of substitutions S192N and G194S+S (B and C). The restored LRR6 allows Asn192\textsubscript{InlA} to form an intramolecular hydrogen-bond to Ser194\textsubscript{InlA}-N while retaining the direct hydrogen bond to Phe17\textsubscript{hEC1}. As a result Asn192 no longer excludes a second water molecule between InlA and hEC1 (see B) weakening the interaction. A shorter hydrogen bond from Glu54\textsubscript{hEC1} to bridging water molecules may indicate tighter interaction with the inserted serine +S. Alternatively, these shorter distances may result from a different crystal packing, as overall tighter binding is not observed thermodynamically. In S192N-G194S+S/hEC1 and G194S+S-Y369S/hEC1, both of which crystallized in space group P1, Table S1, Arg55\textsubscript{hEC1} of hEC1 forms a hydrogen bond to LRR7-10 of a symmetry-related molecule, possibly shifting Glu54\textsubscript{hEC1} towards InlA.
**Supplementary Material**

**Table S1:** Data collection, refinement statistics and structural characteristics of InlA-mutant/hEC1 complexes.

<table>
<thead>
<tr>
<th>Data collection¹</th>
<th>InlA&lt;sub&gt;WT&lt;/sub&gt;/hEC1⁴</th>
<th>InlAY369S/hEC1</th>
<th>InlAG194S-S/hEC1</th>
<th>InlAS192N/hEC1⁴</th>
<th>InlAS192N-Y369S/hEC1</th>
<th>InlAS192N-G194S-S/hEC1</th>
<th>InlAG194S-S-Y369S/hEC1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁,2₁,2₁</td>
</tr>
<tr>
<td><strong>Cell constants</strong></td>
<td>a, b, c (Å)</td>
<td>55.4, 86.8, 110.7</td>
<td>55.3, 84.6, 108.3</td>
<td>55.2, 88.5, 110.5</td>
<td>55.5, 68.9, 132.4</td>
<td>55.4, 87.4, 111.0</td>
<td>44.9, 54.1, 68.6</td>
</tr>
<tr>
<td></td>
<td>α, β, γ (°)</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 90.0</td>
<td>74.7, 80.7, 67.7</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>25-1.8 (1.86-1.80)</td>
<td>40-1.6 (1.63-1.60)</td>
<td>40-2.0 (2.1-2.0)</td>
<td>40-1.7 (1.76-1.70)</td>
<td>50-1.90 (1.97-1.90)</td>
<td>50-1.7 (1.77-1.70)</td>
<td>50-1.8 (1.86-1.80)</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;merge&lt;/sub&gt; (%)</strong></td>
<td>6 (39)</td>
<td>5 (29)</td>
<td>12 (37)</td>
<td>6 (28)</td>
<td>8 (42)</td>
<td>6 (52)</td>
<td>6 (51)</td>
</tr>
<tr>
<td><strong>I / σ&lt;sub&gt;I&lt;/sub&gt;</strong></td>
<td>14 (3)</td>
<td>29 (6)</td>
<td>9 (4)</td>
<td>20 (4)</td>
<td>17 (3)</td>
<td>16 (2)</td>
<td>12 (2)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>92 (84)</td>
<td>99 (95)</td>
<td>96 (92)</td>
<td>99 (95)</td>
<td>97 (86)</td>
<td>91 (77)</td>
<td>94 (75)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>6 (5)</td>
<td>8 (8)</td>
<td>5 (4)</td>
<td>6 (6)</td>
<td>12 (4)</td>
<td>5 (4)</td>
<td>5 (4)</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th><strong>Resolution (Å)</strong></th>
<th>1.8</th>
<th>1.6</th>
<th>2.0</th>
<th>1.7</th>
<th>1.9</th>
<th>1.7</th>
<th>1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. reflections</strong></td>
<td>41618</td>
<td>60684</td>
<td>37285</td>
<td>50104</td>
<td>43193</td>
<td>56061</td>
<td>47450</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt; (%)</strong></td>
<td>17 / 22</td>
<td>15 / 19</td>
<td>19 / 28</td>
<td>16 / 21</td>
<td>15 / 22</td>
<td>17 / 22</td>
<td>16 / 22</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>4577</td>
<td>4645</td>
<td>4524</td>
<td>4704</td>
<td>4630</td>
<td>4797</td>
<td>4691</td>
</tr>
<tr>
<td><strong>Ca&lt;sup&gt;2+&lt;/sup&gt;/Cl&lt;sup&gt;-&lt;/sup&gt;</strong></td>
<td>2 / 3</td>
<td>0 / 0</td>
<td>1 / 1</td>
<td>2 / 4</td>
<td>2 / 2</td>
<td>1 / 1</td>
<td>1 / 1</td>
</tr>
<tr>
<td><strong>Water</strong></td>
<td>541</td>
<td>801</td>
<td>605</td>
<td>697</td>
<td>845</td>
<td>723</td>
<td>783</td>
</tr>
<tr>
<td><strong>B-factors mean</strong></td>
<td>24</td>
<td>11</td>
<td>18</td>
<td>10</td>
<td>18</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td><strong>Wilson</strong></td>
<td>23</td>
<td>16</td>
<td>23</td>
<td>14</td>
<td>21</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td><strong>R.m.s deviations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bond lengths (Å)</strong></td>
<td>0.018</td>
<td>0.022</td>
<td>0.024</td>
<td>0.025</td>
<td>0.022</td>
<td>0.024</td>
<td>0.022</td>
</tr>
<tr>
<td><strong>Bond angles (°)</strong></td>
<td>1.8</td>
<td>1.8</td>
<td>2.1</td>
<td>2.0</td>
<td>1.7</td>
<td>2.0</td>
<td>1.9</td>
</tr>
</tbody>
</table>

**Analysis²**

<table>
<thead>
<tr>
<th>r.m.s.d. (InlA)</th>
<th>-</th>
<th>0.43</th>
<th>0.39</th>
<th>0.59</th>
<th>0.30</th>
<th>0.32</th>
<th>0.24</th>
</tr>
</thead>
<tbody>
<tr>
<td>r.m.s.d. (hEC1)</td>
<td>-</td>
<td>0.47</td>
<td>0.28</td>
<td>0.46</td>
<td>0.24</td>
<td>0.42</td>
<td>0.41</td>
</tr>
<tr>
<td>r.m.s.d. (complex)</td>
<td>-</td>
<td>0.64</td>
<td>0.30</td>
<td>0.61</td>
<td>0.31</td>
<td>0.50</td>
<td>0.57</td>
</tr>
</tbody>
</table>

¹ Numbers in parenthesis correspond to values of the highest resolution shell.
² r.m.s.d. values were calculated using main chain atoms.
³ Schubert, et al. 2002 (10)
⁴ Wollert et al., 2007 (38)