Supporting Information

Yao *et al.* 10.1073/pnas.0900212106

SI Materials and Methods

Plasmids and Reagents. For expression in enteropathogenic *Escherichia coli* (EPEC), cycle-inhibiting factor (Cif) was cloned into pBluescript II SK(−). For recombinant expression in *E. coli*, Cif homolog in *Burkholderia pseudomallei* (CHBP) was constructed into pGEX-6P (Amersham Biosciences). For transfection in mammalian cells, Cif family members were cloned into pCS2(+) with an N-terminal EGFP tag. Truncation and deletion mutants were constructed by standard PCR cloning strategy, and point mutants were generated by the QuikChange site-directed mutagenesis kit (Stratagene). All of the plasmids were verified by DNA sequencing. Cell culture products were from Invitrogen, and all other chemicals were Sigma–Aldrich products unless noted.

Protein Purification and Crystallization. For expression, the BL21 (DE3) strain (Novagen) harboring the CHBP-N48 construct was induced for 8 h at 22 °C in the presence of isopropyl-β-thiogalactopyranoside after A_{600} reached 0.8. The GST fusion protein was purified by glutathione–Sepharose resin, and the GST tag was removed by PreScission protease digestion. The protein was further purified by ion exchange chromatography followed by gel filtration chromatography (Superdex 200; Amersham Biosciences). The SeMet-labeled protein was expressed in the methionine-auxotrophic strain B834 (DE3) (Novagen) (1) and purified the same as that for the native protein. For crystallization, purified CHBP proteins were concentrated to 5 mg/mL in a buffer containing 50 mM Tris·HCl and 0.15 M NaCl (pH 7.9). Crystals of native and SeMet-labeled protein were obtained in 32% PEG 1000, 100 mM sodium cacodylate (pH 6.0), and 5% (vol/vol) glycerol in 7 days by using the hanging-drop vapor diffusion method at 20 °C.

Data Collection, Structural Determination, and Refinement. The anomalous diffraction data (Se-Peak) for the SeMet-labeled CHBP crystal were collected at the KEK radiation facility at a wavelength of 0.9791 Å at 95 K and processed with HKL-2000 (2). The structure was determined and autobuilt by the Se-single-wavelength anomalous diffraction phasing method using the program PHENIX (3), which provided a 75% complete model. The remaining residues of the model were manually built with COOT (4). The structure was further refined to 2.1 Å by using native crystal data in PHENIX and Refmac5.0. The quality of both models was checked with PROCHECK program (5). Statistics of data collection and refinement are listed in Table S1. All structural pictures were drawn by using PyMOL (http://www.delanoscientific.com/).

Cell Culture, Transfection, EPEC Infection, and Cell Cycle Analysis. 293T and HeLa cells obtained from American Type Culture Collection were grown in Dulbecco's modified Eagle's medium (DMEM; HyClone) containing 10% FBS and 2 mM L-glutamine at 37 °C in a 5% CO_2 incubator. Calcium phosphate precipitation was adopted for transfection of 293T cells. Purified recombinant proteins were delivered into HeLa cells by using Profect-P1 protein delivery reagent (Targeting Systems) according to the manufacturer's instructions. HeLa cells were synchronized at the G_1/S boundary by double-thymidine block before protein delivery was performed in serum-free medium for 4 h. FBS was added back to 10%, and cells were further cultured for 14 h until the time of assay.

For EPEC infection, a single EPEC colony was inoculated into 0.5 mL of LB and statically cultured overnight at 37 °C. On the day of infection, bacterial cultures were diluted by 1:30 in DMEM and further cultured at 37 °C with 5% CO_2 for 3 h. Infection was carried out by adding the DMEM culture of EPEC directly to HeLa cells with a multiplicity of infection of 200:1, and the infection was usually continued for 1 h at 37 °C in the 5% CO_2 incubator. Cells were then washed three times with PBS and cultured for 18 h in DMEM supplemented with 10% FBS and 200 μg/mL gentamicin before cell cycle analysis.

To determine the cell cycle distribution, trypsin-treated HeLa cells were fixed and permealized in cold 70% ethanol followed by propidium iodide (PI) staining. DNA content was measured on Cell Lab Quanta SC flow cytometer (Beckman Coulter). To assay transiently transfected cells, trypsin-treated 293T cells were fixed in 3% formaldehyde and permealized in 70% ethanol. Both the PI signal for DNA content and EGFP signal were monitored simultaneously on the flow cytometer, and EGFP-positive cells were gated out and analyzed for PI signal.

E-64 Labeling and Mass Spectrometry Analysis. Twelve micrograms of Cif or CHBP-N48 protein was incubated overnight at 30 °C in a buffer containing 25 mM Tris·HCl (pH 7.6), 150 mM NaCl, and 4 mM DTT, with or without 400 μM E-64. Reactions were stopped by adding SDS loading buffer and subjected to SDS/PAGE. For mass spectrometry analysis, protein gel bands were destained with 25 mM NH_4HCO_3 and 50% methanol and reduced in 10 mM DTT for 40 min at 56 °C followed by alkylation in 55 mM iodoacetamide for 1 h in the dark at 20 °C. Proteins were then in-gel-digested with 10 ng/μL sequencing-grade modified trypsin (Promega) in 50 mM NH_4HCO_3 (pH 7.8) at 37 °C overnight. The resulting tryptic peptides were extracted and analyzed on a LTQ mass spectrometer (Thermo; Fisher Scientific) by nano-HPLC/MS/MS in the data-dependent mode (MS scan mass range was from 350 to 2,000 Da, and the top three most abundant precursor ions from each MS scan were selected for MS/MS scans). The proteins were identified through database searches performed on an in-house Mascot server (version 2.2, Matrix Science). The mass tolerance was 2 Da for precursor ions and 0.5 Da for MS/MS scans. The following variable modifications were chosen for database search: cysteine carbamidomethylation, E-64 labeling of cysteine, and methionine oxidation.

1. Studts JM, Fox BG (1999) Application of fed-batch fermentation to the preparation of isotopically labeled or selenomethionyl-labeled proteins. *Protein Expr Purif* 16:109–119.
2. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
3. Adams PD, *et al.* (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D* 58:1948–1954.
4. Emsley P, Cowtan K (2004) COOT: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
5. Laskowski AR, MacArthur WM, Moss SD, Thornton MJ (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993:283–291.

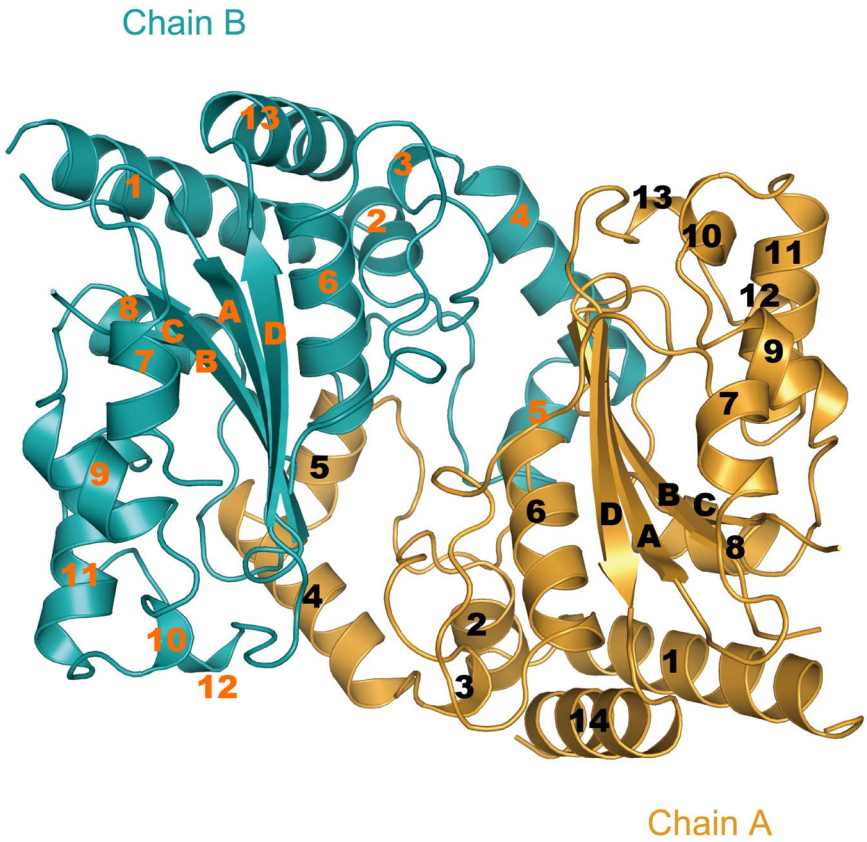

Fig. S1. Diagram of the overall structure of CHBP dimer in one asymmetric unit. Chain A is shown in orange and chain B is in blue. Secondary structures are numbered the same as shown in Fig. 2.

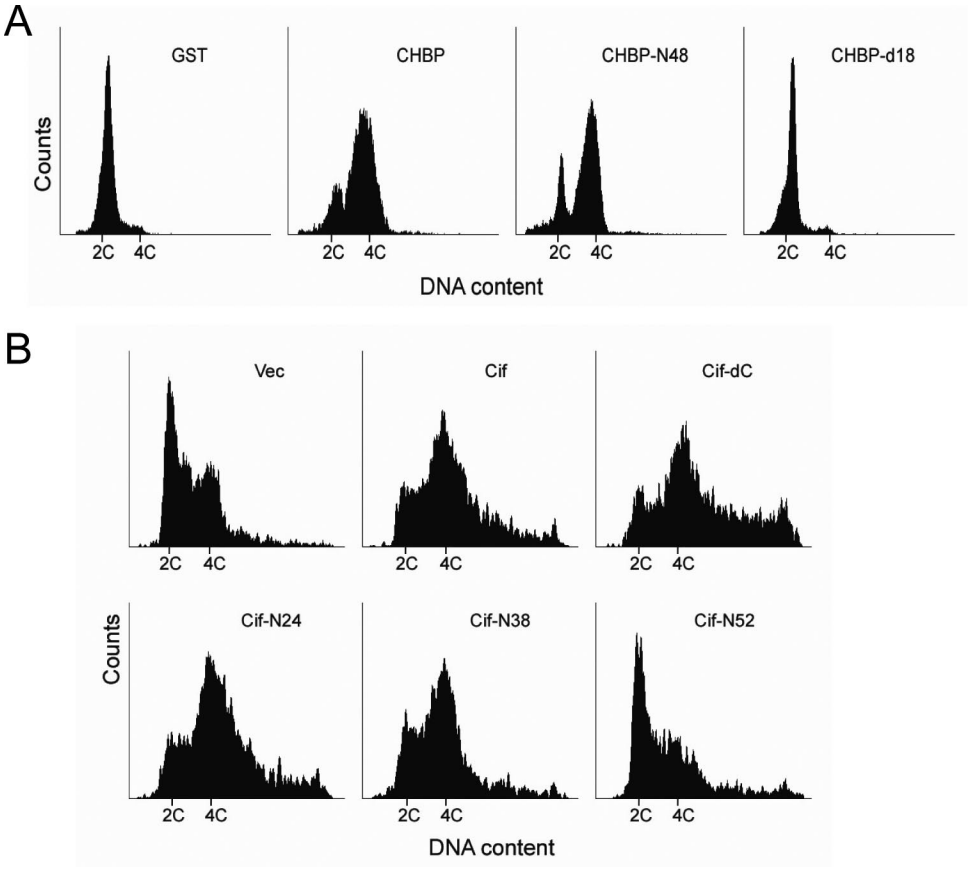

Fig. S2. Effects of truncation and deletion mutations of Cif family members on host cell cycle distribution. (*A*) Effects of CHBP-N48 and CHBP-d18 (deletion of residues 122–139 of CHBP). Purified proteins as indicated were delivered into G_1/S synchronized HeLa cells, and the cell cycle distribution of cells was examined by flow cytometry analysis as shown. (*B*) Cell cycle arrest activity of N-terminal truncation mutants of Cif. DNA contents of 293T cells transfected with indicated Cif constructs were measured. Cif-dC, the natural C-terminal truncation of Cif; Cif-N24, Cif-N38, and Cif-N52, truncations of the N-terminal 23, 37, and 51 residues of Cif, respectively.

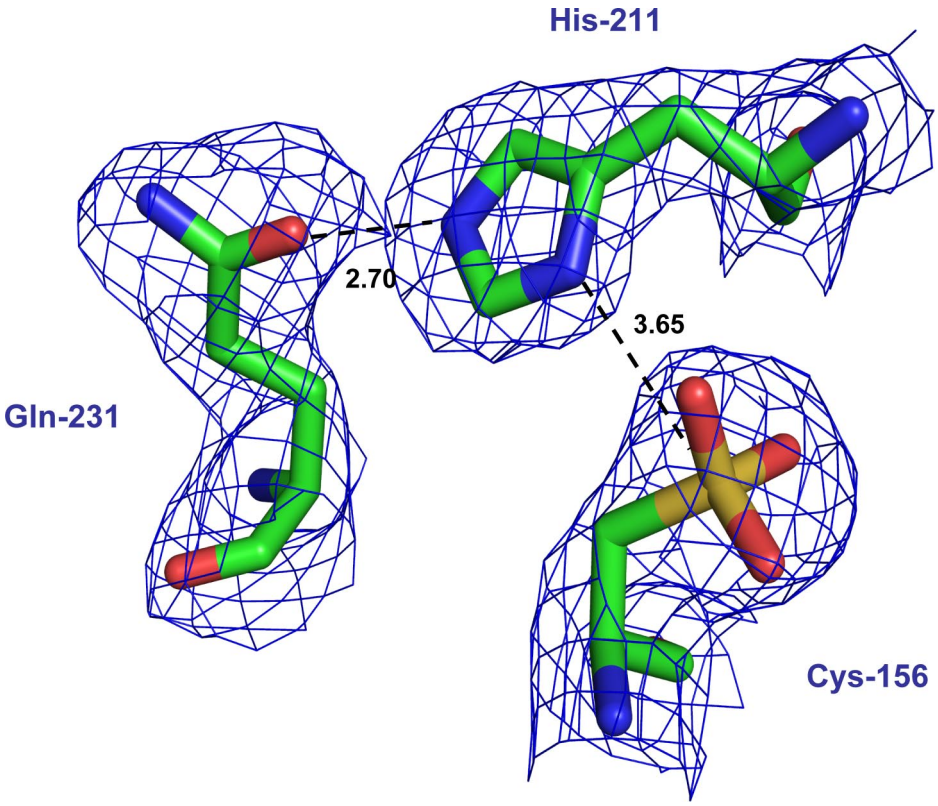

Fig. S3. Electron density map of the putative catalytic residues of CHBP. The oxidized Cys-156, His-211, and Gln-231 are shown in stick representation with σA-weighted 2 $F_o - F_c$ omit electron density map contoured at 1.0 σ.

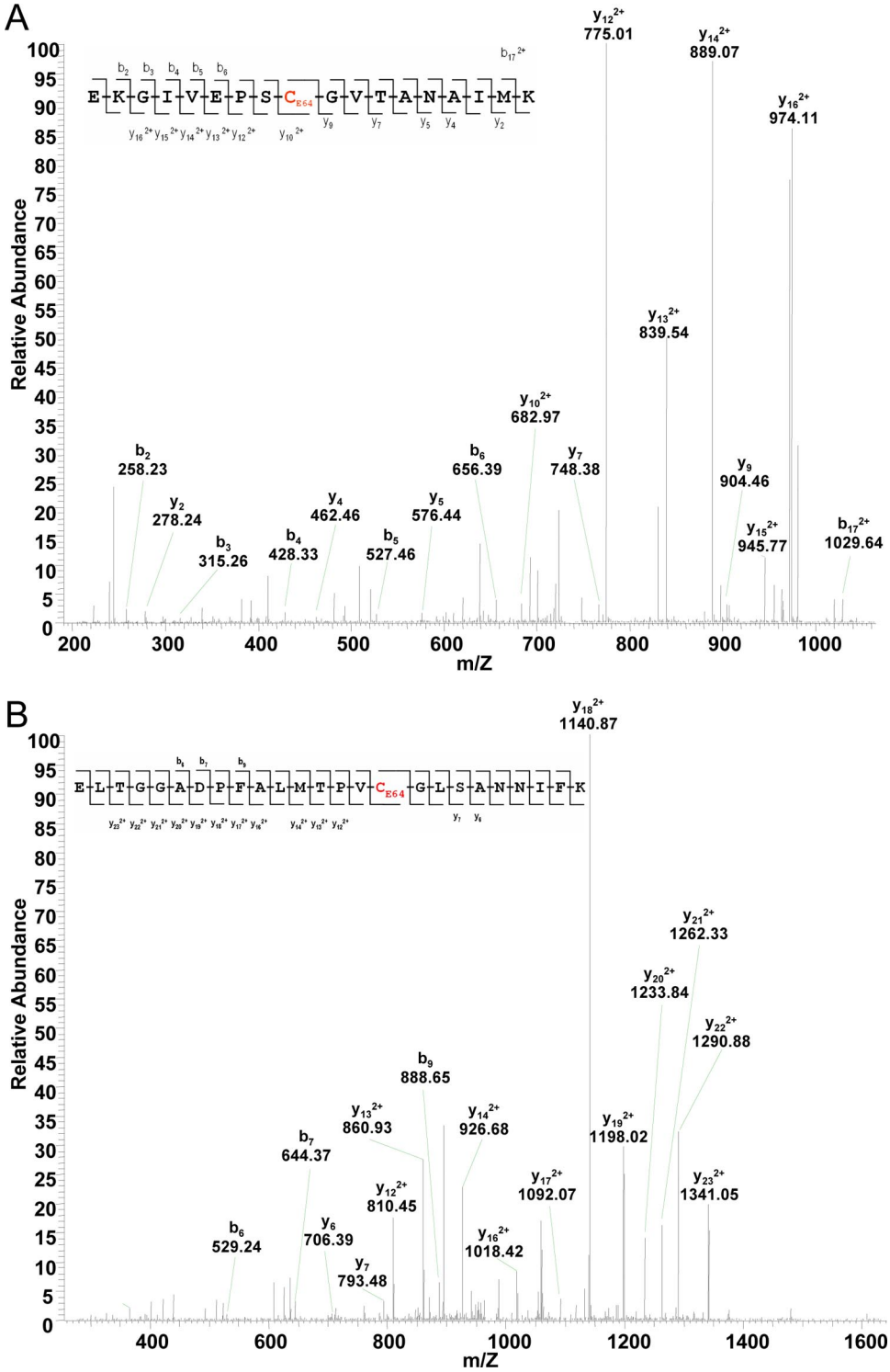

Fig. S4. Mass spectrometry spectra of tryptic peptides of Cif (*A*) and CHBP (*B*) containing E-64-modified cysteine. b and y ions are marked in the spectra. The amino acid sequence of the peptide is shown at *Upper Left* with the modified Cys highlighted in red. The fragmentation patterns that generate the observed b and y ions are illustrated along the peptide sequence.

Table S1. Crystallographic data

Crystal	SeMet	Native
Data collection statistics		
Space group	$P2_12_12_1$	$P2_12_12_1$
Wavelength	0.9791	1.000
a, *b*, *c*, Å	54.34, 78.00, 115.04	54.78, 79.92, 117.42
Resolution range, Å*	50–2.56 (2.65–2.56)	50.0–2.10 (2.18–2.10)
No. of unique reflections	16,544 (1,306)	30,618 (2,990)
Completeness, %	97.9% (78.4%)	99.5% (99.4%)
Redundancy	13.7 (11.2)	6.8 (6.5)
$\langle I/\sigma(I)\rangle$	13.9 (7.1)	18.5 (5.2)
R_{merge}, %	8.8 (30.1)	6.0 (31.8)
Refinement statistics		
R_{work}/R_{free}, %†	20.3/26.7	19.3/25.4
No. of protein atoms	3,857	3,868
No. of water atoms	70	260
Bond length, Å	0.008	0.006
Bond angles, °	1.11	0.94
Ramachandran plot statistics		
Most favored regions, %	98.12	97.92
Additional allowed regions, %	1.88	2.08

*The data for the highest resolution shell are shown in parentheses.
†R_{free} is calculated using 10% of the total number of reflections.

Table S2. Energy terms of dipeptide docking

Dipeptide	Score
ARG-ASN	−39.589
ASN-**ARG**	−39.489
ARG-GLY	−37.087
MET-**ARG**	−35.345
SER-**ARG**	−35.284
GLY-**ARG**	−34.954
ARG-THR	−34.491
ALA-**ARG**	−33.958
ARG-ILE	−33.874
VAL-**ARG**	−32.973
ILE-**ARG**	−32.74
CYS-**ARG**	−32.739
THR-**ARG**	−32.096
ASP-**ARG**	−32.046
ARG-SER	−29.911
ARG-ALA	−29.354
LEU-**ARG**	−29.135
ARG-VAL	−28.396
ASN-GLY	−26.825
GLN-ASN	−26.146
ARG-LEU	−25.306
LEU-ASN	−24.851
GLY-ASN	−23.512
ASN-ASN	−23.493
ARG-CYS	−23.19
ASN-GLN	−23.185
GLU-**ARG**	−23.147
GLN-GLN	−22.724
ASN-THR	−21.822
CYS-ASN	−21.771
GLN-**ARG**	−21.703
VAL-ASN	−21.476
MET-ASN	−21.472
GLN-GLY	−21.469
ASN-SER	−21.389
LEU-GLN	−21.05
ILE-ASN	−20.87
SER-SER	−19.362
ASN-VAL	−19.314
SER-GLN	−19.085
GLN-MET	−19.018
ASN-ILE	−18.763
Dipeptide	Score
GLN-VAL	−18.718
GLN-LEU	−18.436
GLY-GLN	−18.431
ASN-MET	−18.37
GLN-CYS	−18.127
GLN-ALA	−18.06
THR-ASN	−17.901
LYS-ASN	−17.783
CYS-GLN	−17.587
CYS-CYS	−17.581
GLN-ILE	−17.464
ASN-ALA	−17.386
ALA-ASN	−17.37
CYS-GLY	−17.338
GLN-THR	−17.249
GLN-SER	−17.135
MET-GLN	−17.125
ASN-CYS	−17.124
SER-ASN	−17.018
ASP-GLN	−16.642
VAL-GLN	−16.429

Dipeptide	Score
THR-GLN	−16.172
THR-SER	−16.126
ASN-*LYS*	−16.095
ILE-GLN	−16.087
ALA-GLN	−16.039
ARG-GLN	−15.259
GLY-SER	−15.145
TYR-TRP	−14.976
SER-THR	−14.974
ALA-CYS	−14.208
MET-CYS	−14.168
SER-GLY	−14.158
CYS-SER	−14.145
THR-MET	−14.063
GLN-*LYS*	−14.005
MET-SER	−13.909
THR-ALA	−13.823
VAL-SER	−13.761
THR-GLY	−13.705
GLY-CYS	−13.508
SER-VAL	−12.679
Dipeptide	Score
SER-ILE	−12.588
ALA-SER	−12.522
ARG-GLU	−12.411
ASN-LEU	−12.314
LEU-SER	−12.299
LYS-GLN	−12.228
SER-MET	−12.196
LEU-CYS	−12.122
VAL-CYS	−12.092
SER-CYS	−12.047
ILE-CYS	−11.949
THR-CYS	−11.637
THR-LEU	−11.583
THR-THR	−11.431
LYS-SER	−11.181
THR-VAL	−10.864
TYR-VAL	−10.855
CYS-THR	−10.308
SER-LEU	−10.022
TYR-GLN	−9.795
ASN-PHE	−9.622
LYS-CYS	−9.534
LEU-THR	−9.482
SER-ALA	−9.289
CYS-ALA	−9.219
PHE-TYR	−9.142
PHE-GLN	−9.085
VAL-THR	−9.085
TYR-CYS	−8.958
GLN-PRO	−8.444
LEU-GLY	−8.417
ALA-LEU	−8.348
ILE-SER	−8.283
MET-ILE	−8.283
SER-*LYS*	−8.186
THR-ILE	−8.019
TRP-TYR	−8.019
PRO-GLN	−7.896
ILE-THR	−7.731
GLY-THR	−7.6
ALA-THR	−7.387
MET-THR	−7.337
Dipeptide	Score

Dipeptide	Score
TRP-GLY	−7.184
CYS-VAL	−7.067
PRO-ASN	−6.994
PRO-PHE	−6.76
GLY-GLY	−6.759
ALA-GLY	−6.684
GLY-ALA	−6.667
CYS-ILE	−6.635
LEU-MET	−6.496
ASN-TRP	−6.326
ALA-ILE	−6.274
ALA-PHE	−6.056
VAL-GLY	−6.044
GLY-VAL	−5.994
TRP-SER	−5.95
ASP-THR	−5.936
ASP-SER	−5.93
ARG-ASP	−5.837
LYS-THR	−5.826
ALA-MET	−5.712
CYS-LEU	−5.655
PHE-CYS	−5.593
ASP-TRP	−5.525
CYS-MET	−5.463
GLY-TRP	−5.257
PHE-**ARG**	−5.252
MET-GLY	−5.106
VAL-LEU	−5.059
ALA-VAL	−5.047
TYR-MET	−4.988
MET-PHE	−4.828
ARG-TYR	−4.693
CYS-PRO	−4.603
ILE-GLY	−4.552
ILE-MET	−4.43
VAL-MET	−4.413
GLY-MET	−4.408
CYS-PHE	−4.335
LEU-ALA	−4.324
MET-ALA	−4.25
PHE-PHE	−4.219
ASP-ASN	−4.217
PHE-ILE	−4.216
THR-*LYS*	−4.132
SER-TRP	−3.971
ALA-ALA	−3.927
ILE-ILE	−3.827
PHE-TRP	−3.728
MET-MET	−3.698
VAL-ALA	−3.511
ASP-CYS	−3.491
VAL-ILE	−3.487
ILE-ALA	−3.459
ILE-LEU	−3.448
MET-LEU	−3.24
ILE-TYR	−3.153
ILE-VAL	−3.097
LEU-VAL	−2.94
TYR-PHE	−2.929
GLN-GLU	−2.908
GLU-ASN	−2.887
PRO-GLY	−2.883
GLU-GLN	−2.756
PRO-MET	−2.561
PHE-ASN	−2.488

Dipeptide	Score
GLN-PHE	−2.391
VAL-VAL	−2.36
GLY-LEU	−2.252
TRP-VAL	−2.211
TYR-LEU	−2.204
PRO-TRP	−2.189
GLU-*LYS*	−2.14
TYR-ASN	−2.111
PHE-LEU	−2.099
ASP-ILE	−2.034
GLU-CYS	−1.88
PHE-PRO	−1.874
MET-VAL	−1.858
HIS-PHE	−1.849
ASP-LEU	−1.801
LEU-TRP	−1.777
ALA-PRO	−1.772
LEU-LEU	−1.767
SER-PHE	−1.587
GLY-ILE	−1.586
ARG-MET	−1.422
ALA-TRP	−1.407
TYR-**ARG**	−0.994
SER-PRO	−0.957
ASN-PRO	−0.926
PRO-PRO	−0.868
MET-PRO	−0.814
TRP-PHE	−0.78
LEU-PHE	−0.776
PHE-MET	−0.688
PRO-CYS	−0.654
PHE-HIS	−0.637
ASP-ASP	−0.511
ILE-PHE	−0.49
ILE-*LYS*	−0.432
PRO-ALA	−0.404
LEU-ASP	−0.366
ASP-MET	−0.311
GLU-SER	−0.198
TRP-ALA	−0.028
PRO-ILE	0.05
LYS-ALA	0.153
ALA-*LYS*	0.176
GLU-THR	0.401
PRO-TYR	0.424
LEU-ILE	0.446
GLY-*LYS*	0.466
ILE-TRP	0.546
LYS-GLY	0.568
PRO-THR	0.875
CYS-GLU	0.924
CYS-ASP	0.939
ASP-VAL	0.954
ASP-*LYS*	0.999
VAL-*LYS*	1.092
VAL-TRP	1.13
TYR-GLY	1.16
ASN-GLU	1.265
GLY-PRO	1.316
VAL-ASP	1.368
ASP-PRO	1.437
VAL-PRO	1.451
THR-GLU	1.467
GLN-ASP	1.48
ARG-PRO	1.574

Dipeptide	Score
LEU-*LYS*	1.815
ILE-PRO	2.072
MET-TRP	2.328
SER-GLU	2.407
PRO-LEU	2.439
PRO-VAL	2.452
SER-ASP	2.465
ASP-ALA	2.51
MET-ASP	2.638
PHE-*LYS*	2.866
SER-TYR	3.204
HIS-PRO	3.313
THR-PRO	3.321
CYS-*LYS*	3.345
GLU-GLY	3.453
ASN-ASP	3.488
LYS-LEU	3.73
TRP-CYS	3.751
TYR-ALA	3.789
THR-ASP	3.804
PRO-SER	3.9
ASN-TYR	3.97
ASP-GLY	4.614
MET-*LYS*	4.748
LYS-PRO	4.792
PHE-GLY	4.792
ILE-GLU	5.151
TRP-PRO	5.338
LYS-ILE	5.385
GLY-ASP	5.422
PHE-ASP	5.467
TRP-LEU	5.697
LYS-VAL	5.701
GLY-PHE	5.779
GLU-ALA	6.059
PHE-THR	6.265
ALA-ASP	6.282
TYR-ASP	6.312
VAL-PHE	6.318
ASP-GLU	6.418
LYS-MET	6.491
MET-GLU	6.496
ILE-ASP	6.505
LEU-PRO	6.566
VAL-GLU	7.002
PHE-SER	7.005
TRP-ASP	7.34
TRP-ILE	7.573
VAL-TYR	7.77
GLY-GLU	8.135
LYS-*LYS*	8.29
GLU-VAL	8.295
THR-PHE	8.552
GLU-PHE	8.737
GLU-LEU	8.75
GLY-TYR	9.016
ALA-TYR	9.304
GLU-MET	9.469
GLU-ILE	9.483
ARG-*LYS*	9.741
LEU-GLU	10.013
LYS-**ARG**	10.025
ALA-GLU	10.094
PHE-VAL	12.248
GLU-GLU	12.688
PRO-*LYS*	12.765
PHE-GLU	13.225
GLU-ASP	14.108
ASP-TYR	14.773
LYS-ASP	14.992
LYS-GLU	16.133
ASP-PHE	17.445
PRO-ASP	18.051
PRO-GLU	18.363
TYR-ILE	19.07
TYR-GLU	20.391
GLU-TYR	22.169
GLU-PRO	22.402

Left column, dipeptide sequence; right column, energy score of docked dipeptide/CHBP complex. The table includes all the 314 dipeptides, for which binding complex can be generated. Arginine and lysine residues are shown in bold and italics, respectively.