The DLiP library

Kazuyoshi Ikeda, Yugo Shimizu, Yuta Maezawa, Toshiyuki Tashiro, Tomoki Yonezawa, Satoru Kanai, Nobuyoshi Sugaya, Yoshiaki Masuda, Naoko Inoue, Tatsuya Niimi, Keiichi Masuya, Kenji Mizuguchi, Toshio Furuya and Masanori Osawa

The DLiP library is a protein−protein interaction (PPI)-oriented library, which is one of the products of the Japan Agency for Medical Research and Development (AMED) research project. This library was designed based on three-dimensional (3D) structures of PPI interfaces and physicochemical properties of known PPI inhibitors. We selected the compounds from a virtual compound library for a diverse of PPI targets using structure-based drug design (SBDD) approach. Finally, only successfully synthesized compounds were registered in the DLiP library. The procedure of library construction is as follows:

1. Preparation of a virtual compound library

We prepared more than 6 million compound data from the K-Library that is a commercially available virtual compound library provided by Kishida Chemical Co., Ltd. These compounds were desalted and then filtered by the following criteria and filters (i−vi):

i) $450 \leq$ molecular weight ≤ 900

ii) Fsp3 (fraction of sp^3 carbon atoms) ≥ 0.35

iii) Reactive compound filter (Rishton, 1997)

iv) PAINS filter (Baell and Holloway, 2010)

v) Solubility filter (MolProp Toolkit 2.5.4.5, OpenEye)

vi) BlockBuster filter (MolProp Toolkit 2.5.4.5, OpenEye)

After passing through the above filters, Omega (OpenEye) was used to generate the 3D structures of the compounds.

2. Preparation of protein structures

We collected PPI structures from the Protein Data Bank (PDB) (Berman et al., 2000). We focused on the complex structure between a peptide and a protein (receptor-protein). The peptide/protein complex structures were collected if the following criteria were satisfied: the amino acid sequence length of peptide ≤ 20 and of receptor-protein ≥ 50. The X-ray crystal structures with resolution > 2.5 Å and structures resolved by nuclear magnetic resonance (NMR) were discarded. To ensure the diversity of PPI targets, the PPI complex structures were divided into three PPI interface types: two types of secondary structures

(helix and turn-and-strand) and others (motif), according to the structural component of the interface. We selected 117 PPI complex structures including 49, 30, and 38 structures for helix, turn-and-strand, and motif, respectively (Tables 1−3). We used the Alpha Site Finder (MOE, CCG) to detect ligand-binding pockets on PPI interfaces. The pockets used in subsequent docking calculation were assessed using the druggability scores calculated by a machine learning method (Sugaya and Ikeda, 2009).

3. Docking calculation

The OEDocking-FRED (OpenEye) was used for docking calculation against the K-library. To evaluate the docking results, we used the consensus score (Charifson et al., 1999) of the six scoring functions (Chemgauss, PLP, Chemscore, London dG, Affinity dG, and Alpha HB). The top 2% ranked compounds were selected and then clustered using Bayon with GraphSim fingerprint descriptor. The best-scored compound from each cluster was selected, and then evaluated by visual inspection. For helix, strand-and-turn, and motif interface types, approximately 8,000, 6,000, and 8,000 compounds were selected for the subsequent synthetic process, respectively.

4. Synthesis of compounds

We successfully synthesized 4,908, 2,566, and 4,511 compounds for helix, turn-and-strand, and motif, respectively. To expand the chemical space suitable for known PPI inhibitors, we additionally synthesized 2,280 compounds with non-flat (*i.e.*, sphere-like) structures satisfying a criterion of the normalized principal moment of inertia ratios (NPR1 + NPR2) ≥ 1.35, and 949 compounds for adding new structures with novel scaffolds. Finally, a total of 15,214 successfully synthesized compounds were registered in the DLiP library.

Table 1. PDB list of PPI complex structures for helix interface.

PDB ID	peptide chain	peptide name	receptor-protein chain	receptor-protein name
1dva	1dva_X	synthetic peptide	1dva_H	Coagulation factor VII
1hqq	1hqq_G	synthetic peptide	1hqq_C	Streptavidin
1nx1	1nx1_C	Calpastatin	1nx1_A	Calpain small subunit 1
1rbf	1rbf_S	Ribonuclease pancreatic	1rbf_A	Ribonuclease S
1t0j	1t0j_C	Voltage-dependent L-type calcium channel subunit alpha-1C	1t0j_B	Voltage-gated calcium channel subunit beta2a

1t3l	1t3l_B	Voltage-dependent L-type calcium channel subunit alpha-1S	1t3l_A	Voltage-dependent L-type calcium channel subunit beta-2
1y3a	1y3a_F	synthetic peptide	1y3a_B	Guanine nucleotide-binding protein G(i) subunit alpha-1
1yp0	1yp0_B	Nuclear receptor subfamily 0 group B member 2	1yp0_A	Steroidogenic factor 1
2auc	2auc_D	Myosin-A	2auc_C	Myosin A Tail Interacting Protein
2c23	2c23_P	Exoenzyme S	2c23_A	14-3-3 protein beta/alpha
2fx7	2fx7_P	Envelope glycoprotein gp160	2fx7_H	Fab 4E10
2g30	2g30_P	Low density lipoprotein receptor adapter protein 1	2g30_A	AP-2 complex subunit beta
2h1c	2h1c_B	Antitoxin FitA	2h1c_A	Toxin FitB
2nm1	2nm1_B	Synaptotagmin-2	2nm1_A	Botulinum neurotoxin type B
2pmc	2pmc_F	Protein phosphatase CheZ	2pmc_B	Chemotaxis protein CheY
2pv2	2pv2_E	synthetic peptide	2pv2_A	Chaperone SurA
2r7g	2r7g_B	Early E1A protein	2r7g_A	Retinoblastoma-associated protein
2uz6	2uz6_L	Alpha-conotoxin TXIA(A10L)	2uz6_B	Soluble acetylcholine receptor
2vzd	2vzd_C	Paxillin	2vzd_A	Alpha-parvin
2x72	2x72_B	Guanine nucleotide-binding protein G(t) subunit alpha-1	2x72_A	Rhodopsin
2xxm	2xxm_T	synthetic peptide	2xxm_A	Capsid protein P24
2xyi	2xyi_B	Histone H4	2xyi_A	Probable histone-binding protein Caf1
2y9q	2y9q_B	MAP kinase-interacting serine/threonine-protein kinase 1	2y9q_A	Mitogen-activated protein kinase 1
2z5t	2z5t_Q	Cellular tumor antigen p53	2z5t_N	Protein Mdm4
3ax3	3ax3_F	Aldehyde dehydrogenase, mitochondrial	3ax3_E	Mitochondrial import receptor subunit TOM20 homolog
3bev	3bev_C	Hemoglobin subunit alpha-A	3bev_A	MHC class I alpha chain 2
3bl2	3bl2_D	Beclin-1	3bl2_B	Bcl-2 homolog
3d32	3d32_C	synthetic peptide	3d32_A	Gamma-aminobutyric acid receptor-associated protein
3ds4	3ds4_T	synthetic peptide	3ds4_A	HIV-1 capsid protein
3gm1	3gm1_F	Paxillin	3gm1_A	Protein-tyrosine kinase 2-beta
3i1h	3i1h_B	Bcl-2 homologous antagonist/killer	3i1h_A	Bcl-2-related protein A1

3ik5	3ik5_B	T-cell surface glycoprotein CD3 zeta chain	3ik5_A	Protein Nef
3k48	3k48_R	synthetic peptide	3k48_AD	Tumor necrosis factor ligand superfamily member 13
3lnz	3lnz_D	synthetic peptide	3lnz_C	E3 ubiquitin-protein ligase Mdm2
3lrh	3lrh_F	Huntingtin	3lrh_E	Anti-huntingtin VL domain
3p72	3p72_B	synthetic peptide	3p72_A	Platelet glycoprotein Ib alpha chain
3q95	3q95_D	Nuclear receptor coactivator 2	3q95_B	Estrogen receptor
3qis	3qis_B	Sesquipedalian-1	3qis_A	Inositol polyphosphate 5-phosphatase OCRL-1
3rdv	3rdv_E	SLAIN motif-containing protein 2	3rdv_A	CAP-Gly domain-containing linker protein 1
3uvk	3uvk_B	Histone-lysine N-methyltransferase 2D	3uvk_A	WD repeat-containing protein 5
3uym	3uym_D	Type II secretion system protein D	3uym_B	Lipoprotein OutS
3zqg	3zqg_C	Anti-inducer peptide TAP2	3zqg_A	Tetracycline repressor protein class B from transposon Tn10
4ch9	4ch9_C	Serine/threonine-protein kinase WNK4	4ch9_A	Kelch-like protein 3
4fjo	4fjo_B	DNA polymerase kappa	4fjo_A	DNA repair protein REV1
4j2c	4j2c_B	Vacuolar protein sorting-associated protein 51 homolog	4j2c_A	Syntaxin-6
4j8s	4j8s_B	Tristetraprolin	4j8s_A	CCR4-NOT transcription complex subunit 1
4k0u	4k0u_B	Type II secretion system protein D	4k0u_A	Lipoprotein OutS
4nb3	4nb3_D	ATR-interacting protein	4nb3_B	Replication protein A 70 kDa DNA-binding subunit
5dhf	5dhf_D	Serine/threonine-protein kinase RIO2	5dhf_C	Exportin-1

Table 2. PDB list of PPI complex structures for turn-and-strand interface.

PDB ID	peptide chain	peptide name	receptor-protein chain	receptor-protein name

1clv	1clv_I	PROTEIN (ALPHA-AMYLASE INHIBITOR)	1clv_A	PROTEIN (ALPHA-AMYLASE)
1dkd	1dkd_F	12-MER PEPTIDE	1dkd_B	GROEL
1i8i	1i8i_C	EPIDERMAL GROWTH FACTOR RECEPTOR, EGFRVIII PEPTIDE ANTIGEN	1i8i_B	EPIDERMAL GROWTH FACTOR RECEPTOR ANTIBODY MR1SCFV HEAVY CHAIN
1jmt	1jmt_B	SPLICING FACTOR U2AF 65 KDA SUBUNIT	1jmt_A	SPLICING FACTOR U2AF 35 KDA SUBUNIT
1r9n	1r9n_H	Neuropeptide Y	1r9n_D	Dipeptidyl peptidase IV
2i60	2i60_M	[PHE23]M47, SCORPION-TOXIN MIMIC OF CD4	2i60_G	EXTERIOR MEMBRANE GLYCOPROTEIN(GP120)
2peh	2peh_D	Splicing factor 3B subunit 1	2peh_B	Splicing factor 45
2q6g	2q6g_D	Polypeptide chain	2q6g_B	severe acute respiratory syndrome coronavirus (SARS-CoV)
2w2u	2w2u_D	CONSERVED ARCHAEAL PROTEIN	2w2u_B	HYPOTHETICAL P60 KATANIN
2zvl	2zvl_U	DNA polymerase kappa	2zvl_A	Proliferating cell nuclear antigen
3lqj	3lqj_T	Histone H3	3lqj_B	MLL1 PHD3-Bromo
3m53	3m53_B	TAF10 peptide	3m53_A	Histone-lysine N-methyltransferase SETD7
3mmg	3mmg_C	Nuclear inclusion protein B fragment	3mmg_A	Nuclear inclusion protein A
3wnf	3wnf_D	CKIDNC peptide	3wnf_AB	Gag-Pol polyprotein
4bxf	4bxf_C	60S RIBOSOMAL PROTEIN L27A	4bxf_A	BIFUNCTIONAL LYSINE-SPECIFIC DEMETHYLASE AND HISTIDYL-HYDROXYLASE MINA
4dcb	4dcb_F	Plasminogen	4dcb_A	Coagulase/fibrinolysin
4gq6	4gq6_B	Histone-lysine N-methyltransferase MLL	4gq6_A	Menin
4gqb	4gqb_C	Histone H4 peptide	4gqb_A	Protein arginine N-methyltransferase 5
4h3h	4h3h_F	Pol II CTD peptide	4h3h_E	RNA polymerase II subunit A C-terminal domain phosphatase SSU72
4jzw	4jzw_M	CD4-MIMETIC MINIPROTEIN M48U1	4jzw_G	HIV-1 YU2 gp120 glycoprotein

4odm	4odm_K	30S ribosomal protein S2	4odm_D	Peptidyl-prolyl cis-trans isomerase SlyD
4oyk	4oyk_C	Ubiquitin thioesterase otulin	4oyk_A	E3 ubiquitin-protein ligase RNF31
4pr5	4pr5_C	Epstein-Barr nuclear antigen 1	4pr5_A	HLA class I histocompatibility antigen, B-35 alpha chain
4qae	4qae_T	Hepcidin	4qae_E	Neutrophil gelatinase-associated lipocalin
4yiz	4yiz_F	Rhoptry neck protein 2, putative	4yiz_E	Apical membrane antigen AMA1
4z80	4z80_D	Cytoadherence-linked asexual protein	4z80_B	EGF family domain-containing protein
5cx3	5cx3_G	FYVE and coiled-coil domain-containing protein 1	5cx3_C	Microtubule-associated proteins 1A/1B light chain 3A
5fjz	5fjz_R	PROTEIN TRANSPORT PROTEIN DSL1	5fjz_C	COATOMER SUBUNIT DELTA
5fw5	5fw5_C	NON-STRUCTURAL PROTEIN 3	5fw5_A	RAS GTPASE-ACTIVATING PROTEIN-BINDING PROTEIN 1
5jr2	5jr2_G	APYd3 peptide	5jr2_C	Ephrin type-A receptor 4

Table 3. PDB list of PPI complex structures for motif interface.

PDB ID	peptide chain	ELM motif in peptide	receptor-protein chain	receptor-protein name
1d01	1d01_I	LIG_TRAF2_1	1d01_F	TNF receptor-associated factor 2
1ee4 *	1ee4_C	TRG_NLS_MonoExtC_3	1ee4_A	Importin subunit alpha
1jd5	1jd5_B	LIG_BIR_III_3	1jd5_A	Death-associated inhibitor of apoptosis 1
1jpl	1jpl_E	TRG_LysEnd_GGAAcLL_1	1jpl_A	ADP-ribosylation factor-binding protein GGA3
1m7e	1m7e_F	LIG_PTB_Apo_2	1m7e_C	Disabled homolog 2
1nu2	1nu2_B	LIG_PTB_Apo_2	1nu2_A	Disabled homolog 1
1nw9	1nw9_B	LIG_BIR_III_2	1nw9_A	E3 ubiquitin-protein ligase XIAP
1r1s	1r1s_F	LIG_SH2_GRB2	1r1s_E	GRB2-related adaptor protein 2
1tp3	1tp3_B	LIG_PDZ_Class_1	1tp3_A	Disks large homolog 4
1w9e	1w9e_R	LIG_PDZ_Class_2	1w9e_A	Syntenin-1
1w9o	1w9o_T	LIG_PDZ_Class_2	1w9o_A	Syntenin-1

2an6	2an6_F	LIG_SIAH_1	2an6_B	E3 ubiquitin-protein ligase SIAH1A
2cci	2cci_I	LIG_CYCLIN_1	2cci_D	Cyclin-dependent kinase 2
2he2	2he2_B	LIG_PDZ_Class_	2he2_A	Disks large homolog 2
2hqh	2hqh_E	LIG_CAP-Gly_1	2hqh_A	Dynactin subunit 1
2pg1	2pg1_L	LIG_Dynein_DLC8_1	2pg1_B	Dynein light chain 1, cytoplasmic
2wmb	2wmb_I	DOC_CYCLIN_1	2wmb_B	Cyclin-A2
2xl3	2xl3_C	LIG_WD40_WDR5_VDV_1	2xl3_A	WD repeat-containing protein 5
3e87	3e87_C	MOD_PKB_1	3e87_A	RAC-beta serine/threonine-protein kinase
3esk	3esk_B	LIG_TPR	3esk_A	Stress-induced-phosphoprotein 1
3gd1	3gd1_Z	LIG_Clathr_ClatBox_1	3gd1_I	Clathrin heavy chain 1
3gjo	3gjo_F	LIG_SxIP_EBH_1	3gjo_B	Microtubule-associated protein RP/EB family member 1
3ml4	3ml4_E	LIG_PTB_Phospho_1	3ml4_A	Protein Dok-7
3pxe	3pxe_F	LIG_BRCT_BRCA1_1	3pxe_B	Breast cancer type 1 susceptibility protein
3twt	3twt_E	DOC_ANK_TNKS_1	3twt_A	Tankyrase-2
3ua7	3ua7_E	LIG_SH3_2	3ua7_A	Tyrosine-protein kinase Fyn
3ubw	3ubw_P	LIG_14-3-3_1	3ubw_A	14-3-3 protein epsilon
3ukz *	3ukz_C	TRG_NLS_Bipartite_1	3ukz_B	Importin subunit alpha-1
3utm	3utm_C	DOC_ANK_TNKS_1	3utm_A	Tankyrase-1
3v4y	3v4y_H	LIG_PP1	3v4y_G	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit
3zkf	3zkf_L	LIG_Dynein_DLC8_1	3zkf_K	Dynein light chain 1, cytoplasmic
4bld	4bld_F	LIG_SUFU_	4bld_B	Maltose-binding periplasmic protein
4eje	4eje_D	LIG_PTAP_UEV_1	4eje_B	Tumor susceptibility gene 101 protein
4i7b	4i7b_D	DEG_SIAH_1	4i7b_C	E3 ubiquitin-protein ligase SIAH1
4xc2	4xc2_E	LIG_LIR_Gen_1	4xc2_A	Gamma-aminobutyric acid receptor-associated protein
4y32	4y32_C	LIG_14-3-3_3	4y32_A	14-3-3 protein sigma
5azg	5azg_D	LIG_LIR_Gen_1	5azg_B	Protein lgg-1

5d94	5d94_B	LIG_LIR_Gen_1	5d94_A	Microtubule-associated proteins 1A/1B light chain 3B

* The PDB entries have two distantly-located pockets on the peptide/protein interface of the PPI complex structure.