IDENTIFICATION OF CRITICAL HETERODIMER PROTEIN INTERFACE PARAMETERS BY MULTI-DIMENSIONAL SCALING IN EUCLIDIAN SPACE

Cui Zhanhua, Jacob Gah-Kok Gan, Li Lei, Venkatarajan Subramanian Mathura, Meena Kishore Sakharkar, and Pandjassarame Kangueane

[1]*Nanyang Centre for Supercomputing and Visualization, School of Mechanical & Production Engineering, Nanyang Technological University, Singapore,* [2]*Roskamp Institute, 2040 Whitfield Ave, Sarasota FL 34243, USA*

TABLE OF CONTENTS

1. ABSTRACT

Protein subunit dimers are either homodimers (consisting of identical polypeptides) or heterodimers (consisting of different polypeptides). Protein dimers are involved in several cellular processes and an understanding of their molecular principle in complexations (subunit – subunit interaction) is essential. This is generally studied using 3D structures of homodimers and heterodimers determined by X-ray crystallography. However, the current knowledge on subunit interaction is limited due to lack of sufficient 3D dimer structures. It is our interest to study heterodimers using 3D structures to identify interaction parameters that would help in the development of a model to predict heterodimer interaction sites just from protein sequences. The efficiency of such models depends on the weighted contribution of numerous parameters characterizing heterodimer interfaces. Therefore, we studied the salient features of 111 interface parameters in 65 heterodimer structures. In this study, we applied multi-dimensional scaling for dimensionality reduction on these parameters to select the most critical ones that best characterize heterodimer interfaces. The significance of these parameters in subunit interaction is discussed.

2. INTRODUCTION

Protein-protein interactions play a key role in many biological processes such as signal transduction, gene regulation and antibody-antigen recognition (1-2). Therefore, a study on the principles of protein-protein interaction is critical for developing reliable prediction models from sequence data. Current models largely depend on the available knowledge of protein-protein interaction sites (3-5). However, many model parameters have not been fully captured due to limited structural data and lack of rigorous mathematical formulations.

Studies indicate the presence of charge and electrostatic complementation at the protein-protein interfaces (6-7). Formation of hydrogen bonds between subunits plays an important role in the association and stability of protein subunits (8-9). Residue propensity between interior, exterior and interface regions of oligomeric proteins has been examined (10-12). This showed the selective occurrence of non-polar residues in the interior and at interface regions of proteins, while polar (or charged) residues prefer solvent exposed exterior regions. Thus, a number of parameters have been known to characterize

protein-protein interfaces. Nonetheless, it is extremely difficult to capture all the non-linear dependencies of such parameters.

A number of methods have been used to identify interface parameters in oligomeric complexes. These methods utilize conserved residues at interface (13), surface patches (14), sequence features (15-17), atomic contact vectors (ACV) (18), topological entities (19), neural network trained sets (20), interface energy landscapes (21), and support vector machines (SVM) (22). However, these methods lack sufficient parameters for model development and are often less conclusive in prediction. Here, we analyze 111 interface parameters in 65 heterodimer structures to select the most critical ones in subunit interactions using a multidimensional procedure described elsewhere (35).

3. MATERIALS AND METHODS

3.1. Creation of a heterodimer structural dataset
We created a dataset of 65 high resolution (≤ 3Å) heterodimer structures determined by X-ray crystallography for this analysis (Table 1). These structural data were obtained from the protein databank (PDB). The dataset was selected such that each polypeptide in these heterodimers is at least 50 residues long.

3.2. Interface parameters
Each of the 65 heterodimer interfaces was studied using 111 parameters and the corresponding values were determined. The parameter list is given in Table 2. Consequently, a 65 X 111 matrix was generated for the 65 heterodimers.

3.3. Parameter normalization
Each parameter value was normalized such that standard deviation is equal to one and the average is equal to zero. The standard deviation was calculated using STDEVP function in Microsoft Excel. The normalization ensures that all parameters are expressed as dimensionless numbers. The normalized parameter value is represented by S using α (parameter index whose range is from 1 to 111), i (heterodimer interface index whose range is from 1 to 65), P (parameter value), n (number of heterodimers i.e. 65 in number), $\overline{P_\alpha}$ (parameter mean) and $\sigma_{P\alpha}$ (standard deviation). By definition, S is given as

$$s_\alpha(i) = \frac{(P_\alpha(i) - \overline{P_\alpha})}{\sigma_{P\alpha}} \quad (1)$$

where,

$$\sigma_{P\alpha} = \sqrt{\frac{n\sum_{i=1}^{n} P_\alpha(i)^2 - (\sum_{i=1}^{n} P_\alpha(i))^2}{n^2}} = \sqrt{\frac{65\sum_{i=1}^{65} P_\alpha(i)^2 - (\sum_{i=1}^{65} P_\alpha(i))^2}{4225}} \quad (2)$$

This procedure generated a 65 x 111 matrix containing normalized parameter values.

3.4. Representation of parameters
There are 65 heterodimer structures used in this analysis and each dimer interface (i) is represented as a

vector $\overrightarrow{S(i)}$ in 111 dimensional 'continuous space', where the components $S_\alpha(i)$ are the normalized parameter values. The scalar product Q_{ij} between two vectors $S_\alpha(i)$ and $S_\alpha(i)$, where j is another index for a heterodimer interface, is given by

$$Q_{ij} = \overrightarrow{S(i)} \bullet \overrightarrow{S(j)} = \sum_{\alpha=1}^{111} s_\alpha(i) \cdot s_\alpha(j) \quad (3)$$

The 65 x 65 matrix Q is positive symmetric consisting of the scalar products of the parameter vectors S(i) and S(j) , where i = 1 to 65 and j = 1 to 65.

3.5. Calculation of eigenvectors and eigenvalues
The symbolic eigenvectors (E) of a square matrix Q and eigenvalues (λ) of Q are computed, respectively, using the MATLAB command E = eig(Q). The eigenvalues of Q are the zeros of the characteristic polynomial of Q. As Q is of order 65, we will have 65 eigenvectors and eigenvalues λ and the smallest eigenvalue λ_{65} is near zero due to normalization of the parameters. The eigenvalues and their corresponding eigenvectors are indexed in decreasing order of eigenvalues.

3.6. Selection of interface descriptors
The distribution of the eigenvalues of the Q matrix (Figure 1), containing the scalar products between all pairs of the 111 dimensional heterodimer vectors, rapidly decreases from the largest value λ_1 to λ_{65}. The rapid decrease of the eigenvalues derived from the 111 physical – chemical parameters shows a large anisotropy of the distribution of the parameter values. This anisotropy is a consequence of the large redundancy in the sets of parameter values. This suggests that the number of parameters can be reduced while retaining approximately the same distribution of heterodimers in the property space. We found that the eigenvalues rapidly decrease within the first six largest eigenvalues.

3.7. Calculation of distances in the parameter space
If μ represents the index of eigenvalue and eigenvector, each heterodimer can be represented as a vector in a six-dimensional Euclidean space with each dimension perpendicular to each other. The co-ordinates of the ith heterodimer can be written as:

$$\sqrt{\lambda_{\mu=1}} E_i^\mu \quad (4)$$

where μ varies from 1 to 6.

The distance between the ith and j^{th} heterodimer interface is given by

$$d_{ij} = \sqrt{\sum_{\mu=1}^{n} (\sqrt{\lambda_\mu} E_i^\mu - \sqrt{\lambda_\mu} E_j^\mu)^2} \quad (5)$$

where n is 6.

Distances computed between heterodimers in the six dimensional Eigen sub-space constitute the parameter

Table 1. Dataset used in this analysis

PDB Code	Resolution (Å)	Protein name	PDB Code	Resolution (Å)	Protein name
1A14	2.5	Single chain antibody – Neuraminidase	1DVF	1.9	Fv d1.3 – Fv e5.2
1A2Y	1.5	Monoclonal antibody D1.3 – lyzozyme	1MKX	2.2	Alpha-thrombin – Prethrombin-2
1AR1	2.7	Antibody fv fragment -- Cytochrome C oxidase	1AHW	3.0	Immunoglobulin Fab 5g9 –Tissue factor
1JRH	2.8	Antibody A 6 – Interferon-gamma receptor alpha chain	1GPW	2.4	Hisf protein – Amidotransferase hish
1KB5	2.5	Antibody desire-1 – Kb5-C20 t-cell antigen receptor	3FRU	2.2	Neonatal Fc receptor - - Beta-2-microglobulin
1KIQ	1.9	Antibody D1.3 – Lysozyme	1B8M	2.8	Brain derived neurotrophic factor –Neurotrophin-4
1NMB	2.5	Fab nc10 – N9 neuraminidase	1BJ3	2.6	Coagulation factor ix-binding protein A – Coagulation factor ix-binding protein B
1NSN	2.9	Igg fab - - Staphylococcal nuclease	1BLX	1.9	Cyclin-dependent kinase 6 – P19ink4d
1OSP	2.0	Fab 184.1 – Outer surface protein A	1BND	2.3	Brain derived neurotrophic factor – Neurotrophin 3
1WEJ	1.8	E8 antibody – Cytochrome C	1DKF	2.5	Retinoid x receptor-alpha -- Retinoic acid receptor-alpha
2JEL	2.5	Jel42 Fab fragment – Histidine-containing protein	1DOA	2.6	Gtp-binding protein -- Gdp-dissociation inhibitor 1
1BJ1	2.4	Fab fragment –Vascular endothelial growth factor	1DS6	2.4	Ras-related c3 botulinum toxin substrate 2 -- Rho gdp-dissociation inhibitor 2
1BVK	2.7	Hulys11 – Lysozyme	1E96	2.4	Ras-related c3 botulinum toxin substrate 1 -- Neutrophil cytosol factor 2
1DZB	2.0	Scfv fragment 1f9 –Turkey egg-white lysozyme C	1GZQ	2.3	T-cell surface glycoprotein cd1b -- B2-microglobulin
1JTP	1.9	Single-domain antibody – Lysozyme C	1H32	1.5	Diheme cytochrome C -- Cytochrome C
1MEL	2.5	Vh single-domain antibody – Lysozyme	1SPP	2.4	Major seminal plasma glycoprotein Psp-I--Major seminal plasma glycoprotein Psp-II
1NMC	2.5	Single chain antibody –Neuraminidase	1HE8	3.0	Phosphatidylinositol 3-kinase catalytic subunit,gamma isoform --Transforming protein p21/h-ras-1
1JTT	2.1	Vh single-domain antibody – Lysozyme	1VRK	1.9	RS20 – Calmodulin
1AVG	2.6	Thrombin –Ttriabin	1F7Z	1.6	Trypsin II, anionic – Pancreatic trypsin inhibitor
1AZZ	2.3	Collagenase – Ecotin	1FSS	3.0	Acetylcholinesterase – Fasciculin II
1BGX	2.3	Taq DNA polymerase – TP7 MAB	1FY8	1.7	Trypsin II, anionic – Pancreatic trypsin inhibitor
1A4Y	2.0	Angiogenin – Ribonuclease inhibitor	1JLT	1.4	Phospholipase A2 – Phospholipase A2 inhibitor
1BI8	2.8	Cyclin-dependent kinase 6 – Cyclin-dependent kinase inhibitor	1JTD	2.3	Tem-1 beta-lactamase – Beta-lactamase inhibitor protein II
1BTH	2.3	Thrombin – Bovine pancreatic trypsin inhibitor	1K9O	2.3	Trypsin ii anionic – Alaserpin
1EAI	2.4	Elastase – Chymotrypsin/elastase isoinhibitor 1	1SGP	1.4	Streptomyces griseus proteinase B – Turkey ovomucoid inhibitor
1TBR	2.6	Thrombin – Rhodniin	1SLU	1.8	Anionic trypsin – Trypsin inhibitor
1TFX	2.6	Trypsin – Tissue factor pathway inhibitor	1SMP	2.3	Serratia metallo proteinase – Erwinia chrysanthemi inhibitor
1AY7	1.7	Guanyl-specific ribonuclease SA – Barstar	1TAW	1.8	Trypsin -- Protease inhibitor domain of alzheimer's amyloid beta-protein precursor
1C9P	2.8	Trypsin – Bdellastasin	1UDI	2.7	Uracil-Dna glycosylase -- Uracil-Dna glycosylase inhibitor protein
1CT4	1.6	Proteinase B – Ovomucoid inhibitor	1VIW	3.0	Alpha-amylase -- Alpha-amylase-inhibitor
1DFJ	2.5	Ribonuclease A – Ribonuclease inhibitor	2SGP	1.8	Proteinase B -- Ovomucoid inhibitor
1DHK	1.9	Porcine pancreatic alpha-amylase – Bean lectin-like inhibitor	3TGK	1.7	Trypsin ii, anionic -- Pancreatic trypsin inhibitor
1F34	2.5	Pepsin A – Major pepsin inhibitor PI-3			

distance matrix (PDM). Small distances values between two heterodimers indicate that they are similar in all of the 111 physical and chemical parameters.

3.8. Calculation of correlation coefficient

Pearson's correlation coefficient between pairs of parameter values (x_i, y_i) is calculated using the correlation function (CORRCOEF) in MATLAB.

4. RESULTS

4.1. Quantitative descriptors for heterodimer interface parameters in six dimensions

We used 65 high resolution heterodimer structures (Table 1) to derive a comprehensive list of 111 physical/chemical parameters for heterodimer interfaces

(Table 2). Each heterodimer was represented as a vector in the 111-dimensional space of normalized parameters with mean value of zero and standard deviation 1. Our multi-dimensional scaling approach reveals the high redundancy of the parameter values. The computational approach and justification for reduction to a lower dimensional space follows closely the practice of embedding in distance geometry and it is easy to eliminate redundant variables when describing complex phenomenon in molecular recognition. The distribution of eigenvalues decreases rapidly (Figure 1). This is due to large redundancy in the parameter set. This suggests that the number of parameters can be reduced while retaining approximately the same distribution of heterodimers in the parameter space. The eigenvalues rapidly decrease within the first six largest eigenvalues. We compared distances in the original

Heterodimer protein interface

Table 2. List of 111 heterodimer interface parameters

No.	Parameter names	No.	Parameter names
1	Interface area (Å^2)[†]	57	Interface TYR
2	Interface non-polar area (Å^2)	58	Interface VAL
3	Interface polar area (Å^2)	59	Hydrophobic residues per 100 interface residues
4	RMSD[‡]	60	Polar residues per 100 interface residues
5	Interface salt bridge[#]	61	S-containing residues per 100 interface residues
6	Interface salt bridges per 100 Å^2 interface area	62	Charged residues per 100 interface residues
7	Interface salt bridges per interface residue	63	Negative charge residues per 100 interface residues
8	Gap volume (Å^3)[‡]	64	Positive charge residues per 100 interface residues
9	Gap index	65	Interface ALA per 100 interface residues
10	Interface helix residues	66	Interface ARG per 100 interface residues
11	Interface strand residues	67	Interface ASN per 100 interface residues
12	Interface loop residues	68	Interface ASP per 100 interface residues
13	Ratio of interface helix residues to interface residues	69	Interface CYS per 100 interface residues
14	Ratio of interface strand residues to interface residues	70	Interface GLN per 100 interface residues
15	Ratio of interface loop residues to interface residues	71	Interface GLU per 100 interface residues
16	Interface H-bonds[††]	72	Interface GLY per 100 interface residues
17	Interface H-bonds between main chain – main chain	73	Interface HIS per 100 interface residues
18	Interface H-bonds between main chain – side chain	74	Interface ILE per 100 interface residues
19	Interface H-bonds between side chain – main chain	75	Interface LEU per 100 interface residues
20	Interface H bonds between side chain – side chain	76	Interface LYS per 100 interface residues
21	Interface H-bonds per 100 Å^2 interface area	77	Interface MET per 100 interface residues
22	Interface H-bonds between main chain - main chain per 100 Å^2 interface area	78	Interface PHE per 100 interface residues
23	Interface H-bonds between main chain - side chain per 100 Å^2 interface area	79	Interface PRO per 100 interface residues
24	Interface H-bonds between side chain - main chain per 100 Å^2 interface area	80	Interface SER per 100 interface residues
25	Interface H-bonds between side chain - side chain per 100 Å^2 interface area	81	Interface THR per 100 interface residues
26	Interface H-bonds per interface residue	82	Interface TRP per 100 interface residues
27	Interface H-bonds between main chain – main chain per interface residue	83	Interface TYR per 100 interface residues
28	Interface H-bonds between main chain – side chain per interface residue	84	Interface VAL per 100 interface residues
29	Interface H-bonds between side chain – side chain per interface residue	85	Interface residues per 100 complex residues
30	Interface H-bonds between side chain – main chain per interface residue	86	Interface hydrophobic residues per 100 complex residues
31	Interface hydrophobicity[§]	87	Interface polar residues per 100 complex residues
32	Interface residues	88	Interface S-containing residues per 100 complex residues
33	Interface hydrophobic residues	89	Interface charged residues per 100 complex residues
34	Interface polar residues	90	Interface negative charge residues per 100 complex residues
35	Interface sulphur containing residues	91	Interface positive charge residues per 100 complex residues
36	Interface charged residues	92	Interface ALA per 100 complex residues
37	Interface negative charge residues	93	Interface ARG per 100 complex residues
38	Interface positive charge residues	94	Interface ASN per 100 complex residues
39	Interface ALA	95	Interface ASP per 100 enzyme residues
40	Interface ARG	96	Interface CYS per 100 complex residues
41	Interface ASN	97	Interface GLN per 100 complex residues
42	Interface ASP	98	Interface GLU per 100 complex residues
43	Interface CYS	99	Interface GLY per 100 complex residues
44	Interface GLN	100	Interface HIS per 100 complex residues
45	Interface GLU	101	Interface ILE per 100 complex residues
46	Interface GLY	102	Interface LEU per 100 complex residues
47	Interface HIS	103	Interface LYS per 100 complex residues
48	Interface ILE	104	Interface MET per 100 complex residues
49	Interface LEU	105	Interface PHE per 100 complex residues
50	Interface LYS	106	Interface PRO per 100 complex residues
51	Interface MET	107	Interface SER per 100 complex residues
52	Interface PHE	108	Interface THR per 100 complex residues
53	Interface PRO	109	Interface TRP per 100 complex residues
54	Interface SER	110	Interface TYR per 100 complex residues
55	Interface THR	111	Interface VAL per 100 complex residues
56	Interface TRP		

† Interface area is calculated by NACESS (29). ‡ Gap volume between subunits and RMSD of interface are calculated SURFBET (30). # The number of salt bridges formed between subunits are counted by WHATIF (31). †† The number of H-bonds formed between subunits is counted by HBPLUS (32). § Interface hydrophobicity $= \dfrac{\sum_{i=1}^{N}(HV)}{N}$ (33), N is the number of interface residues, and HV is the hydrophobicity scale for residues (34).

parameter space with those regenerated from a subset of n eigenvectors, varying n systematically from 2 to 65 (Figure 2). The correlation coefficient between the original and regenerated distances was more than 95% for $n = 6$, and approaches 1 very rapidly. We therefore chose the first six eigenvalues and eigenvectors to calculate the six dimensional descriptors of the heterodimer interfaces. The individual distances in the original parameter space and in the sub-space using the first six eigenvectors were highly correlated (Figure 3). The correlation coefficient between the distances was 0.96.

4.2. Selection of critical interface parameters from highly correlated descriptors

We used the first six highly correlated descriptors (dimension 65 X 65) and normalized parameter values (dimension 65 X 111) to calculate the correlation coefficients between the selected descriptors (E$_1$ to E$_6$) and
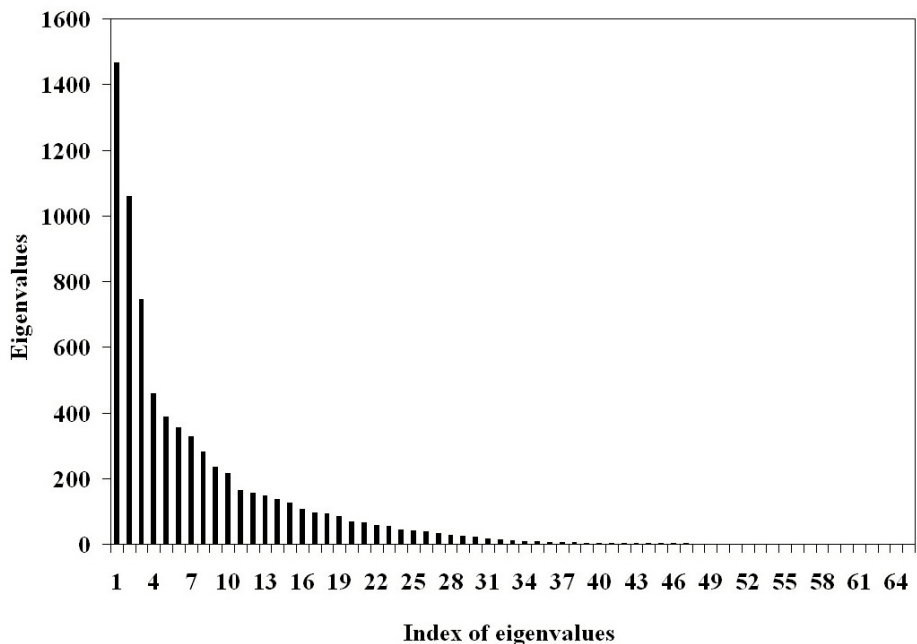
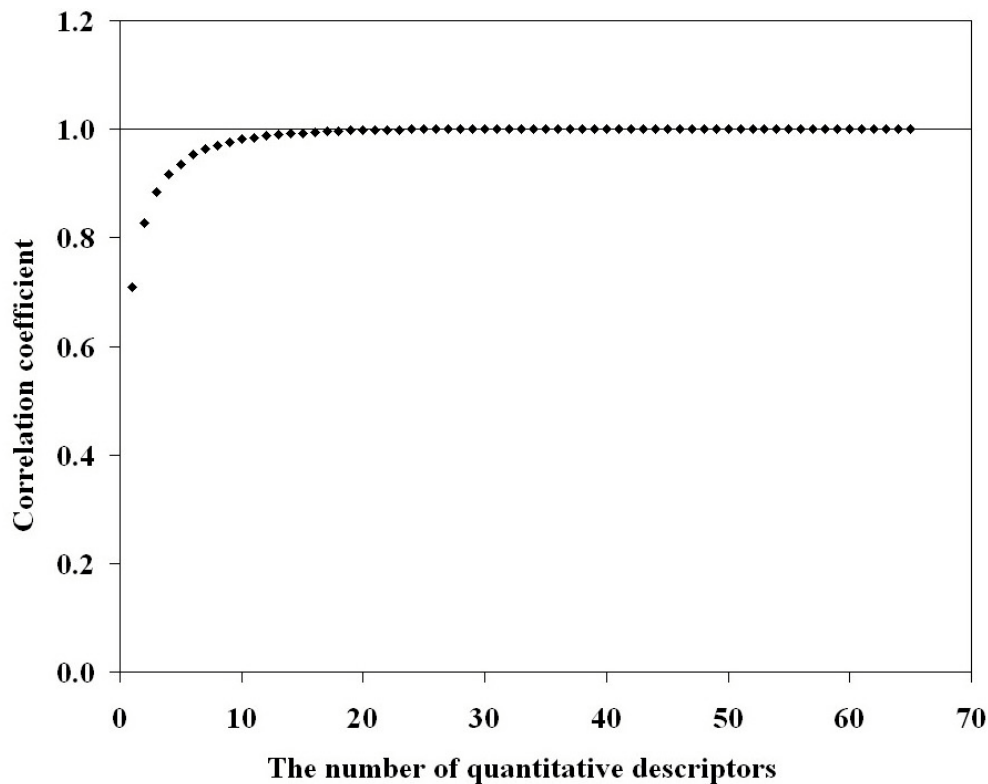**Figure 1.** Distribution of eigenvalues

Figure 2. Correlation coefficients for parameter distances.

the original normalized parameter values. This operation generated a matrix (dimension 6 X 111) containing correlations between the six highly correlated descriptors and the normalized parameter values. We then used this matrix to select the most significant interface parameters using the calculated correlation coefficients (Table 3). This further enabled us to select the most significant parameters that best describe a heterodimer protein interface (Table 3). We then used the parameter values for these six parameters to calculate its distances from the rest of 111 original

Table 3. Critical parameters for heterodimer interfaces

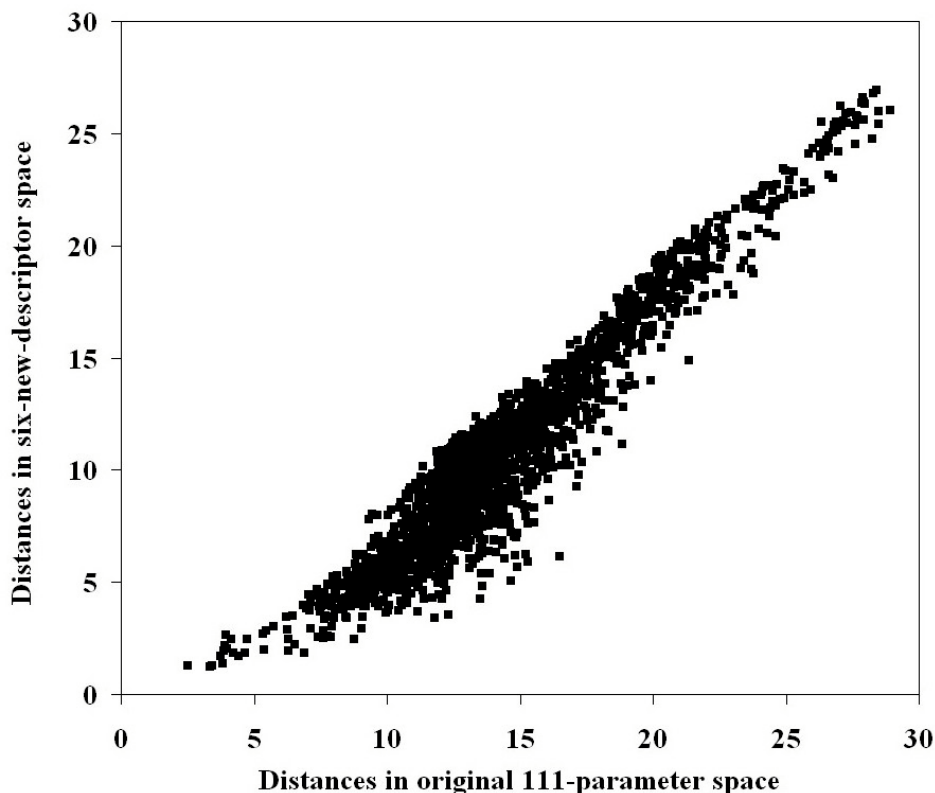| Top six eigenvectors	Parameter name	Correlation coefficients between the distances of top descriptors (E1 to E6) and the original normalized parameter values	Correlation coefficients between the distances of selected parameter values and the original normalized parameter values
E_1	Interface H-bond	0.63	0.61
E_2	Interface TRP	0.51	0.40
E_3	Interface residues	0.79	0.40
E_4	Interface hydrophobicity	0.70	0.36
E_5	Interface coil	-0.77	0.32
E_6	Interface MET	-0.36	0.32

Figure 3. The distance between six selected descriptors and 111 parameter space is given. The linear correlation coefficient is 0.95.

parameter values. The distances were then used to calculate the correlation coefficients. These values suggest that these six parameters have different weights in heterodimer subunit interactions. Data shows that the H-bonds have the highest weight among the six parameters listed in Table 3. We also calculated the individual distances between the original parameter values and the six selected parameter values (Figure 4). The correlation coefficient between these distances was found to be 0.7.

5. DISCUSSION

Heterodimer protein interaction is a common phenomenon in cellular regulation and signaling. This occurs by a huge combination of physical-chemical parameters that characterize their interacting surfaces. The multi dimensional scaling method applied in this study helps to reduce a large pool of interface parameters to a small set of six quantitative descriptors of heterodimer

interfaces. Here, we show that the six parameters (Figure 3) were sufficient to reproduce the distances in the complete parameter space (Figure 4). The most significant parameters that are found to reproduce the original parameter set are given in Table 3. They are dominated by (1) interface H-bonds, (2) interface tryptophan, (3) interface residues, (4) interface hydrophobicity, (5) interface coils and (6) interface methionine. It should be noted that several linear combinations of parameter values represent a descriptors and it is often difficult to further refine or simplify such non-linearity. The goal here is to identify the most critical parameters that represent hetero dimer interfaces. In general it is difficult to decide *a priori* which of the many parameters dominate at the interface. Our quantitative descriptors represent a precise spatial relation of all hetero dimers with respect to the 111 physical-chemical parameters. This enabled us to identify the most critical parameters and these parameters are further discussed below.
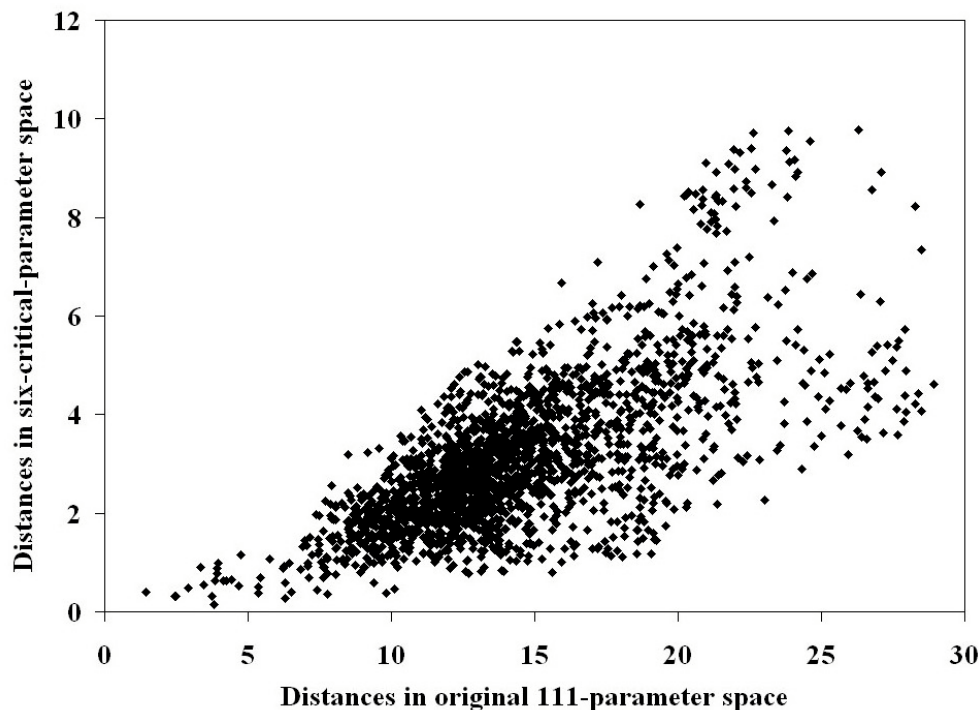
Figure 4. The distance between six selected parameters and 111 parameter space is given. The linear correlation coefficient is 0.71.

5.1. Interface H-bonds

Intermolecular hydrogen bonds between subunits are important in the association and stability of hetero-dimers (26). This analysis suggests that interface H-bonds have a good correlation co-efficient of r = 0.61 with the distances of other interface parameters.

5.2. Interface tryptophan and methionine

Aromatic and aliphatic residues have greater propensity at the protein-protein interfaces (27-28). As given in Table 3, the correlation coefficients of interface tryptophan and methionine with the numerical descriptors (E_2 and E_6) are 0.51 and -0.36. This relation is weak. In fact, E_2 is a descriptor that describes a combination of aliphatic residues and E_6 is a descriptor that describes a combination of aromatic residues. In this study, tryptophan and methionine residues were chosen as prominent parameters because of their high correlation coefficients compared to other members of aliphatic or aromatic residue groups.

5.3. Interface residues and interface hydrophobicity

Hydrophobicity plays an important role in protein association (23-24). Thus, interface hydrophobicity was among the prominent parameters for heterodimer interaction. The number of interface residues relates to interface area. Stronger protein subunit associations were generally associated with larger interface areas (12). These parameters are shown to be used in the prediction of heterodimer interaction sites by surface patch analysis

(14). The method detects the most possible interaction sites by the incorporation of this parameter.

5.4. Interface loop residues

It has been shown that secondary structural elements at the interface play an important role in heterodimer protein assembly (12). Studies also suggest that protein active sites might appear in coiled regions (25). Thus, interface loop residues have critical role in heterodimer interaction.

6. CONCLUSIONS

A large number of structurally important physical - chemical parameters characterize heterodimer interfaces and each of these parameters contributes differently to the stability of a heterodimer interface. A weighted value was assigned to each parameter to indicate the differential contribution. Here, we apply a mathematical procedure to determine the most critical parameters that describe a heterodimer interface. The six critical interface parameters discussed here are based on the selected 65 hetero-dimer structures. The multi-dimensional scaling procedure suggests that the six critical parameters effectively replace the original 111 parameter set. These findings are of critical importance in the understanding and development of prediction models for heterodimer interfaces.

7. ACKNOWLEDGEMENTS

8. REFERENCES

1. Loregian, A., Mardsen, H. S. & Palu, G: Protein–protein interactions as targets for antiviral chemotherapy. *Rev. Med. Virol.* 12, 239-262 (2002)

2. Shulman-Peleg, A., Nussinov, R. & Wolfson, H. J: Recognition of Functional sites in proteins structures. *J. Mol. Biol.* 339, 607-633 (2004)

3. Halperin, I., Ma, B., Wolfson, H. J. & Nussinov, R: Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47, 409-443 (2002)

4. Xu, D., Tsai, C. J. & Nussinov, R: Hydrogen bonds and salt bridges across protein–protein interfaces. *Protein Eng.* 10, 999-1012 (1997)

5. Xu, D., Lin, S. L. & Nussinov, R: Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J. Mol. Biol.* 265, 68-84 (1997)

6. Airlie J. McCoy, V. Chandana Epa & Peter M. Colman: Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.* 268, 570-584 (1997)

7. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R: Conservation of polar residues as hot spots at protein interfaces. *Proteins*, 39, 331-342 (2000)

8. Fernandez, A. & Scheraga, H. A: Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *PNAS*, 100(1), 113-118 (2003)

9. Meyer, M., Wilson, P. & Schomburg, D: Hydrogen Bonding and Molecular Surface Shape Complementarity as a Basis for Protein Docking. *J. Mol. Biol.* 264, 199-210 (1996)

10. Argos, P: An investigation of protein subunit and domain interfaces. *Protein Eng.* 2, 101-113 (1988)

11. Janin, J., Miller, S. & Chothia, C: Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* 204, 155-164 (1988)

12. Jones, S. & Thornton, J. M: Protein-protein interactions: A review of protein dimer structures. *Prog. Biophys. Mol. Biol.* 63, 31-65 (1995)

13. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R: Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *PNAS* 100(10), 5772-5777 (2003)

14. Jones, S. & Thornton, J. M: Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.* 272,133-143 (1997)

15. Panchenko, A. R., Kondrashov, F. & Bryant, S: Prediction of functional sites by analysis of sequence and structure conservation. *Protein Sci.* 13, 884-892 (2004)

16. Zhou, H. X. & Shan, Y. B: Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* 44, 336-343 (2001)

17. Ofran, Y. & Rost, B: Predicted protein–protein interaction sites from local sequence information. *FEBS Letters* 544, 236-239 (2003)

18. Mintseris, J. & Weng, Z. P: Atomic contact vectors in protein-protein recognition. *Proteins* 53, 629-639 (2003)

19. Chou, K. C. & Cai, Y. D: A novel approach to predict active sites of enzyme molecules. *Proteins* 55, 77-82 (2004)

20. Fariselli, P., Pazos, F., Valencia, A. & Casadio, R: Prediction of protein-protein interaction sites in heterocomplexes with neural networks. *Eur. J. Biochem.* 269, 1356-1361 (2002)

21. Recio, J. F., Totrov, M. & Abagyan, R: Identification of Protein–Protein Interaction Sites from Docking Energy Landscapes. *J. Mol. Biol.* 335, 843-865 (2004)

22. Koike, A. & Takagi, T: Prediction of protein–protein interaction sites using support vector machines. *Protein Eng.* 17, 165-173 (2004)

23. Archakov1, A. I., Govorun1, V. M., Dubanov1, A. V., Lewi, P. & Janssen, P: Protein-protein interactions as a target for drugs in proteomics. *Proteomics* 3, 380–391 (2003)

24. Larsen, T. A., Olson, A. J. & Goodsell, D. D: Morphology of protein–protein interfaces. *Structure* 6(4), 421-427 (1998)

25. Stephens, D. J. & Banting, G: Direct Interaction of the *trans*-Golgi Network Membrane Protein, TGN38, with the F-actin Binding Protein, Neurabin. *J. Biol. Chem.* 274(42), 30080–30086 (1999)

26. Conte, L. Lo., Hhothia, C & Janin, J: The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* 285, 2177-2198 (1999)

27. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J: A Dissection of Specific and Non-specific Protein–Protein Interfaces. *J. Mol. Biol.* 336, 943-955 (2004)

28. Bahadur, R. P., Chakrabarti, P., Rodier. F. & Janin, J: Dissecting subunit interfaces in homodimeric proteins. *Proteins* 53, 708-719 (2003)

29. Hubbard, S. J. & Thornton, J. M: Biochemistry and Molecular Biology, University College, London (1993)

30. McDonald, I. K. & Thornton, J. M: Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238, 777-793 (1994)

31. Laskowski, R.A: SURFNET: a program for visualizing molecular surfaces, cavities, interfaces. *J. Mol. Graph.* 13(5), 323-330 (1995)

32. Rodriguez, R., Chinea, G., Lopez, T. P. & Vriend, G: Homology modeling, model and software evaluation: three related resources. *CABIOS* 14**,** 523-528 (1998)

33. Jones, S. & Thornton, J. M: Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* 272,121-132 (1997)

34. Radzicka, A. & Wolfenden, R: Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* 27, 1664-1670 (1988)

35. Venkatarajan, M. S. & Braun W: New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J Mol Model* 7, 445-453 (2001)

Send correspondence to: Pandjassarame Kangueane Ph.D, N3-2c-113b, School of Mechanical and Production Engineering, Nanyang Technological University, Singapore - 639798, Tel: +65 6790 5836, Fax: +65 6774 4340, E-mail: MPandjassarame@ntu.edu.sg