Amino Acids (2013) 44:1307–1316
DOI 10.1007/s00726-013-1465-5

ORIGINAL ARTICLE

S-linked protein homocysteinylation: identifying targets based on structural, physicochemical and protein–protein interactions of homocysteinylated proteins

Yumnam Silla · Elayanambi Sundaramoorthy · Puneet Talwar · Shantanu Sengupta

Received: 12 October 2012 / Accepted: 23 January 2013 / Published online: 12 February 2013

Abstract An elevated level of homocysteine, a thiol-containing amino acid is associated with a wide spectrum of disease conditions. A majority (>80 %) of the circulating homocysteine exist in protein-bound form. Homocysteine can bind to free cysteine residues in the protein or could cleave accessible cysteine disulfide bonds via thiol disulfide exchange reaction. Binding of homocysteine to proteins could potentially alter the structure and/or function of the protein. To date only 21 proteins have been experimentally shown to bind homocysteine. In this study we attempted to identify other proteins that could potentially bind to homocysteine based on the criteria that such proteins will have significant 3D structural homology with the proteins that have been experimentally validated and have solvent accessible cysteine residues either with high dihedral strain energy (for cysteine–cysteine disulfide bonds) or low pKa (for free cysteine residues). This analysis led us to the identification of 78 such proteins of which 68 proteins had 154 solvent accessible disulfide cysteine pairs with high dihedral strain energy and 10 proteins had free cysteine residues with low pKa that could potentially bind to homocysteine. Further, protein–protein interaction network was built to identify the interacting partners of these putative homocysteine binding proteins. We found that the 21 experimentally validated proteins had 174 interacting partners while the 78 proteins identified in our analysis had 445 first interacting partners. These proteins are mainly involved in biological activities such as complement and coagulation pathway, focal adhesion, ECM-receptor, ErbB signalling and cancer pathways, etc. paralleling the disease-specific attributes associated with hyperhomocysteinemia.

Keywords Homocysteine · Homologous structure · Disulfide · Dihedral strain energy · pKa · Protein–protein interaction

Abbreviations

Hcy	Homocysteine
DSE	Dihedral strain energy
DSSP	Dictionary of secondary structure prediction
PPI	Protein–protein interaction
DAVID	Database for annotation, visualization and integrated discovery
GO	Gene ontology
CC	Cellular compartment
BP	Biological process
MF	Molecular function

Introduction

Homocysteine (Hcy), a thiol-containing amino acid, is a key intermediate in the methionine metabolic pathway. An elevated level of Hcy is implicated as an independent risk factor for cardiovascular disease (Eikelboom et al. 1999) and is also associated with a wide spectrum of other disease

Electronic supplementary material The online version of this article (doi:10.1007/s00726-013-1465-5) contains supplementary material, which is available to authorized users.

Y. Silla · E. Sundaramoorthy · P. Talwar · S. Sengupta (✉)
CSIR-Institute of Genomics and Integrative Biology,
Mall Road, Delhi 110007, India
e-mail: shantanus@igib.res.in

Present Address:
E. Sundaramoorthy
Department of Oncology, Medical Research Council,
Cancer Cell Unit Hutchison/MRC Research Centre,
University of Cambridge, Hills Road, Cambridge CB2 0XZ, UK

conditions like neural tube defects (Kirke et al. 1997), schizophrenia (Regland et al. 1995), stroke (Perry et al. 1995), placental abnormalities (de Vries et al. 1997), diabetes (Meigs et al. 2001) and Alzheimer's disease (Seshadri et al. 2002), etc. Although elevated levels of Hcy have been associated with various clinical conditions, the mechanism underlying Hcy-induced pathogenesis is not clearly understood. In circulation about 70–80 % of Hcy is bound to proteins, while about 20 % exists as low-molecular-weight oxidized disulfides (homocystine and homocysteine-cysteine mixed disulfide) while free reduced Hcy accounts for only a small fraction (<2 %) of total circulating Hcy (Andersson et al. 1995; Mansoor et al. 1992; Sengupta et al. 2001a, b). Since a majority of the circulating Hcy is bound to proteins, one of the mechanism put forth to explain the cytotoxicity of Hcy is its ability to bind free protein cysteine residues or cleave accessible cysteine disulfide bonds (S-homocysteinylation) thereby altering the structure and/or function of the protein (Sengupta et al. 2001a, b; Jacobsen et al. 2005). Several studies have highlighted the importance of cysteine residues in maintaining protein structure and function. Cysteines are involved in protein folding and stability by forming disulfide bonds, are redox active, coordinate with metal ions, susceptible to reversible oxidation and are highly conserved across the species.

The mechanism of protein S-homocysteinylation has been elucidated (Sengupta et al. 2001a, b). Although several proteins could potentially bind to Hcy, to date only 21 proteins (referred as 'seed proteins') have been experimentally proven to bind with this amino acid (Sundaramoorthy et al. 2008; Minagawa et al. 2010; Tang et al. 2011). All these proteins have been identified based on a candidate protein approach. However, since Hcy is associated with diverse diseases involving various tissues, it can be perceived that if protein S-homocysteinylation is responsible for the etiology of even a fraction of these diseases, then there should be several other proteins that could bind Hcy. Using a global approach we had earlier identified 50 proteins that could bind Hcy based on high cysteine content (>5 %) of the protein and the availability of solvent accessible disulfides with high dihedral strain energy (DSE) or free cysteine with low pKa in the protein (Sundaramoorthy et al. 2008). These parameters were selected based on previous studies that have postulated the three parameters viz solvent accessibility, dihedral strain energy of a cysteine disulfide (Gilbert 1995; Schmidt et al. 2006) and pKa of free cysteine (Sengupta et al. 2001a, b) for predicting the reactivity of cysteines. However, our earlier study was based on the proteins that had high cysteine content (Sundaramoorthy et al. 2008) which may not be mandatory for a protein to be able to bind Hcy. Further, homocysteinylation of proteins might modulate the interaction of these proteins with their interacting partners resulting in altered signalling cascades, the implications of which are not thoroughly understood for hyperhomocysteinemia.

Further, it has been reported earlier that proteins with two or more disulfide bonds have structural similarity even with <25 % sequence identity (Chuang et al. 2003). In another study it has been shown that about 54 % of all the disulfide bonds compared between homologous pairs are conserved based on pairwise comparisons of homologous protein domains and the extent of conservation of disulfide in homologous proteins is unrelated to the overall sequence identity between homologues (Thangudu et al. 2008). In this study we identified several unique proteins that are structurally similar to those that have been already proven to bind Hcy and meet the physicochemical criteria for reactivity of cysteine residues. We also provide a list of proteins whose interaction with the potential homocysteinylated proteins could be perturbed as a consequence of their S-homocysteinylation.

Materials and methods

Homology structure search

To identify the proteins that have structural similarity with the 21 proteins that have been experimentally proven to bind Hcy, the PDB ids of these proteins were searched using iPBA (Gelly et al. 2011). The iPBA provides structurally related protein using SCOP version 1.75 (Murzin et al. 1995) as the structure data set. Based on the Protein Block (PB) alignment score the top 100 hits are reported and values >1.5 are considered to be structurally related with high confidence. Hence in our analysis we have considered all the human proteins that have score greater than 1.5. To quantify the PB sequence alignment score, iPBA further provides a GDT_PB score which is similar to the Global distance test total score (GDT_TS) (Zemla 2003) for the top hundred hits.

Solvent accessibility, dihedral strain energy and pKa calculation

The resulting 3D structure homologous protein data set were analysed for solvent accessibility, dihedral strain energy for disulfide bonded cysteine residues and pKa for free cysteine. The WHAT IF (version 10.1) web server was used for determining surface accessibility and disulfide-specific dihedral angles (Rodriguez et al. 1998). WHAT IF server utilizes the algorithm developed by Kabsch and Sander, implemented in the dictionary of secondary structure prediction (DSSP) (Kabsch and Sander 1983). The 'Cysteine

bridge torsion angles' tool from WHAT IF web server was used for determining disulfide-specific dihedral angles (Rodriguez et al. 1998). Then the DSE of each disulfide bond was predicted from the magnitude of five CHI angles ($\chi1$, $\chi2$, $\chi3$, $\chi1'$ and $\chi2'$) using AMBER force-field formula (Weiner et al. 1984) as mentioned earlier (Sundaramoorthy et al. 2008). The solvent accessibility of proteins having DSE greater than 17.8 kJ/mol was assessed using "Secondary structure, symmetry and accessibility" tool from WHAT IF server. The solvent accessibility of cysteine was also confirmed using PDBe PISA (Protein Interfaces, Surfaces and Assemblies) (Krissinel and Henrick 2007) from EBI web server available at http://www.PDBe.org/pisa. The structure-based pKa values of solvent accessible free cysteine residues were obtained using PROPKA (version 3.1) web server (Søndergaard et al. 2011) based on empirical approach and also using semi-empirical based pKa prediction approach using H++ (version 3.0) web server (Anandakrishnan et al. 2012).

Protein–protein interaction (PPI) network

To identify the direct interacting partners of the seed proteins and putative Hcy binding proteins identified in this study, we built two protein–protein interaction networks using APID2NET (Hernandez-Toro et al. 2007) plugin in Cytoscape version 2.6.0 (Shannon et al. 2003). The APID2NET (APID) server collects literature-curated protein interaction information from various databases such as BIND, BioGrid, DIP, HPRD, IncAct and MINT and then retrieve protein–protein interaction network of user-provided gene or protein list. UniProt ids of the 21 experimentally validated homocysteinylated proteins and 78 structurally homologous proteins (putative target proteins) were provided separately as input ids in APID server to build the interaction networks. For creating each network, in the first step we considered only those interactions that have been validated in at least two experiments using the appropriate APID retrieval search filter option (Hernandez-Toro et al. 2007) in order to increase stringency and minimize false-positive interactions resulting in a PPI network. However, for proteins that did not have any interacting partners validated by two experiments, the interacting partners of all the single nodes were retrieved by taking the criteria of one experimental validation resulting in another PPI network. Three Cytoscape tools viz Advanced Network Merge, Network Analyzer (Assenov et al. 2008), and Cerebral (Barsky et al. 2007) were then applied for modelling each PPI network. The Advanced Network Merge was used to model a final PPI network by taking union of both the PPI networks. Duplicated edges and self loops were also removed using "Advance Network Analyzer" (Assenov et al. 2008). Isolated nodes were also manually removed from the final PPI network. Protein sub-cellular localization information retrieved from uniprot database (Wu et al. 2006) (accessed on 12th December, 2012) was imported as a node attribute. Then, "Cerebral" (Barsky et al. 2007) was applied to the final PPI network to layout all nodes according to their sub-cellular localization such as extracellular, cell membrane, cytoplasm, nucleus and ER.

GO annotation, biological pathway and disease association analysis

To assess the PPI networks in the context of gene ontology (GO), pathway enrichment and disease association, the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tool (version 6.7) (da Huang et al. 2009a, b) was used. The functional annotation clustering of significantly over-represented GO term [cellular compartment (CC)] was retrieved by using GOTERM_CC_ALL option in DAVID functional annotation tool. Similar analysis was also done using GOTERM_MF_ALL and GOTERM_BP_ALL option for molecular function (MF) and biological process (BP). The default setting parameters and multiple corrections by the Benjamini method were used to determine the significant enrichment score of ≥ 1.3 (da Huang et al. 2009a, b). From DAVID Functional annotation tool 'KEGG pathway database' option was then used for identifying the enriched pathways and 'GENETIC_ASSOCIATION_DB_DISEASE database' option was used for identifying disease association enrichment. The threshold value of Enrichment Score was set at 1.3 while other settings were left unchanged (da Huang et al. 2009a, b).

Results

The objective of this study was to identify proteins that could potentially bind with Hcy based on physicochemical properties of proteins that are structurally homologous to the proteins that have been experimentally identified to bind Hcy. To date, there are 21 proteins that have been reported to bind Hcy (Sundaramoorthy et al. 2008; Minagawa et al. 2010; Tang et al. 2011) of which the PDB structures of 19 are available. These 19 proteins were subjected to structural analysis and protein–protein interaction network analysis. The entire work flow is summarized in Fig. 1.

3D structural homologs of S-homocysteinylated proteins

We identified 3D structural homologs of experimentally identified Hcy interacting proteins using iPBA (Gelly et al. 2011). iPBA server revealed 286 PDB ids that corresponds

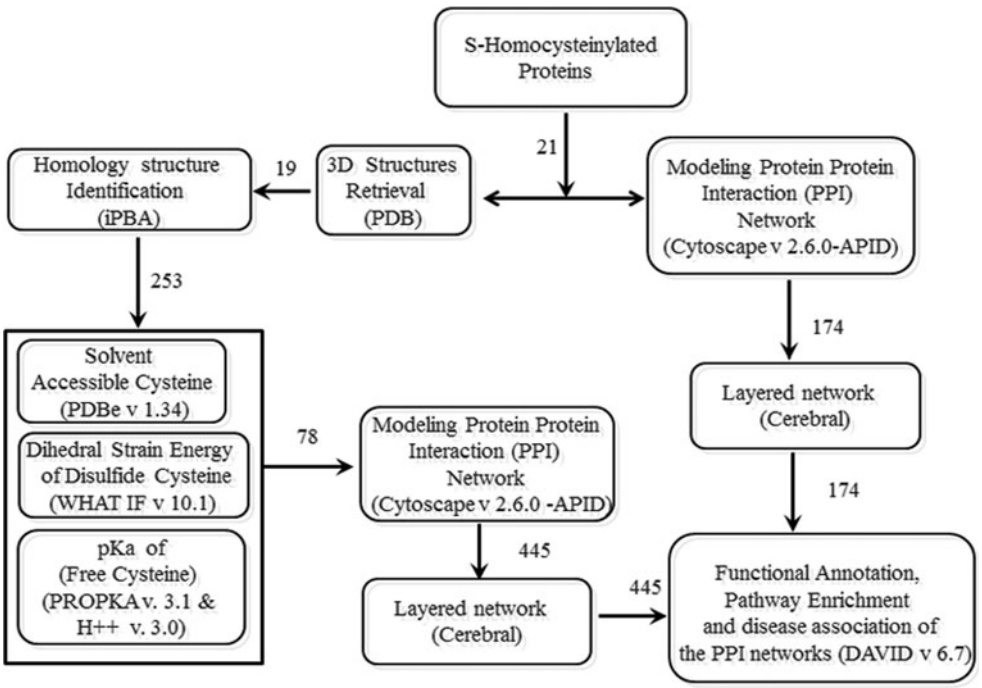

Fig. 1 *Flow charts* describing the entire work flow

to 253 proteins that were structurally similar to the queried proteins by considering PB score greater than 1.5 (Table S1). Among the proteins that were experimentally shown to bind with Hcy TPA (tissue plasminogen activator protein; PDB id: 1a5h) and PROC (activator protein C; PDB id: 1aut) have the maximum number of 3D homologous structures as derived from iPBA (Table S1).

Structural analysis of cysteine-specific physicochemical parameters: solvent accessible, dihedral strain energy or pKa

Hcy either binds to protein free cysteine residues (with low pKa) or could cleave accessible cysteine disulfide bonds (with high DSE) via thiol disulfide exchange reaction (Sengupta et al. 2001a, b; Sundaramoorthy et al. 2008). Thus, in the present study we analysed the three parameters, solvent accessibility of disulfide bonded cysteines, its DSE and pKa of free cysteine residues in the 253 homologous protein structures retrieved using iPBA server. In this data set we found a total of 357 disulfide bond pairs in 112 proteins. The DSE of all the disulfide bonds were calculated as described in "Materials and methods". The dihedral strain energy distribution of 357 disulfide bonds and their energy are shown in Fig. 2 and supplementary Table S2, respectively. The disulfide bond pair formed between the Cys 629- and Cys 660 of MASP2 (Mannan-binding lectin serine protease 2), had the highest DSE value (72.33 kJ mol^{-1}). We chose an empirical cut off of DSE greater than 17.8 kJ/mol for identifying reactive cysteines, based on earlier reports (Ahamed et al. 2006; Sundaramoorthy et al. 2008). Ahamed et al. had experimentally shown that the disulfides found between cysteine residues 186 and 209 in the coagulant protein tissue factor (1AHW) act as a control switch that dictated functional fate of the protein between coagulation and signalling. This particular disulfide has been considered to have a high DSE of 17.8 kJ/mol. Thus, we considered this value as an empirical cut-off in our dataset for identifying reactive disulfides (Ahamed et al. 2006). Using this cut-off in the present study we identified 158 disulfide bonds of which 154 disulfide bonds in 68 proteins were surface accessible (Table S2). Out of these six were found to be interface residues as revealed by PDBe PISA web tool (Krissinel and Henrick 2007).

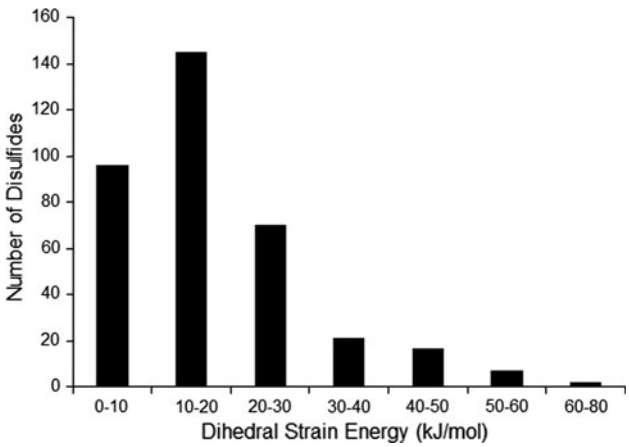

Fig. 2 The distribution of dihedral strain energy (DSE) values of 357 disulfide bonded cysteine residues in 112 proteins

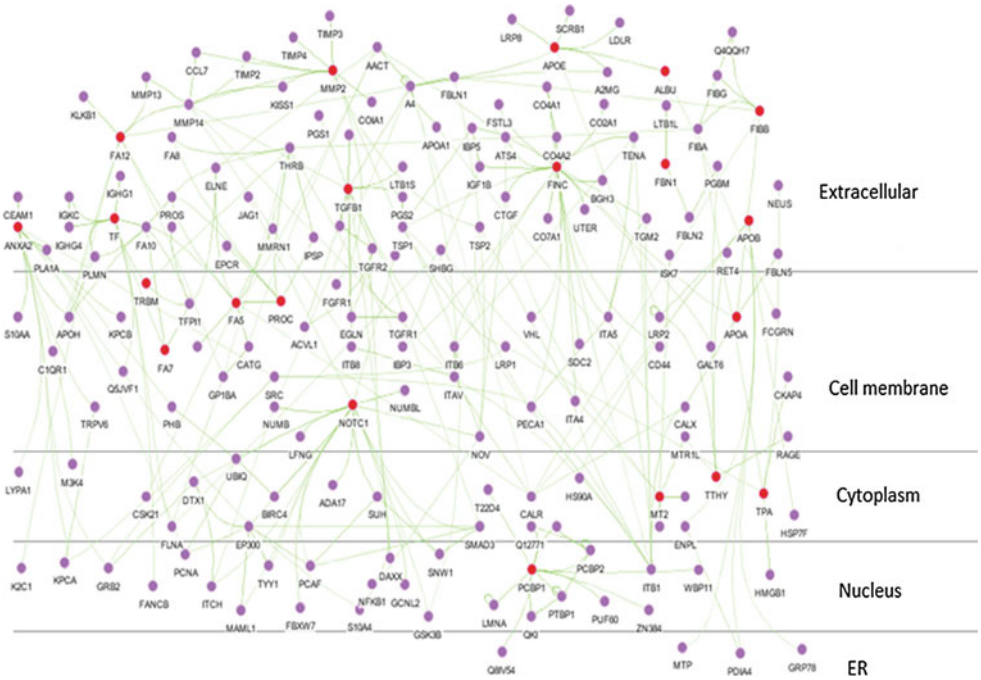

Fig. 3 Layered PPI network: 21 proteins and their interacting partners ($n = 174$) represented as *red* and *violet nodes* respectively (color figure online)

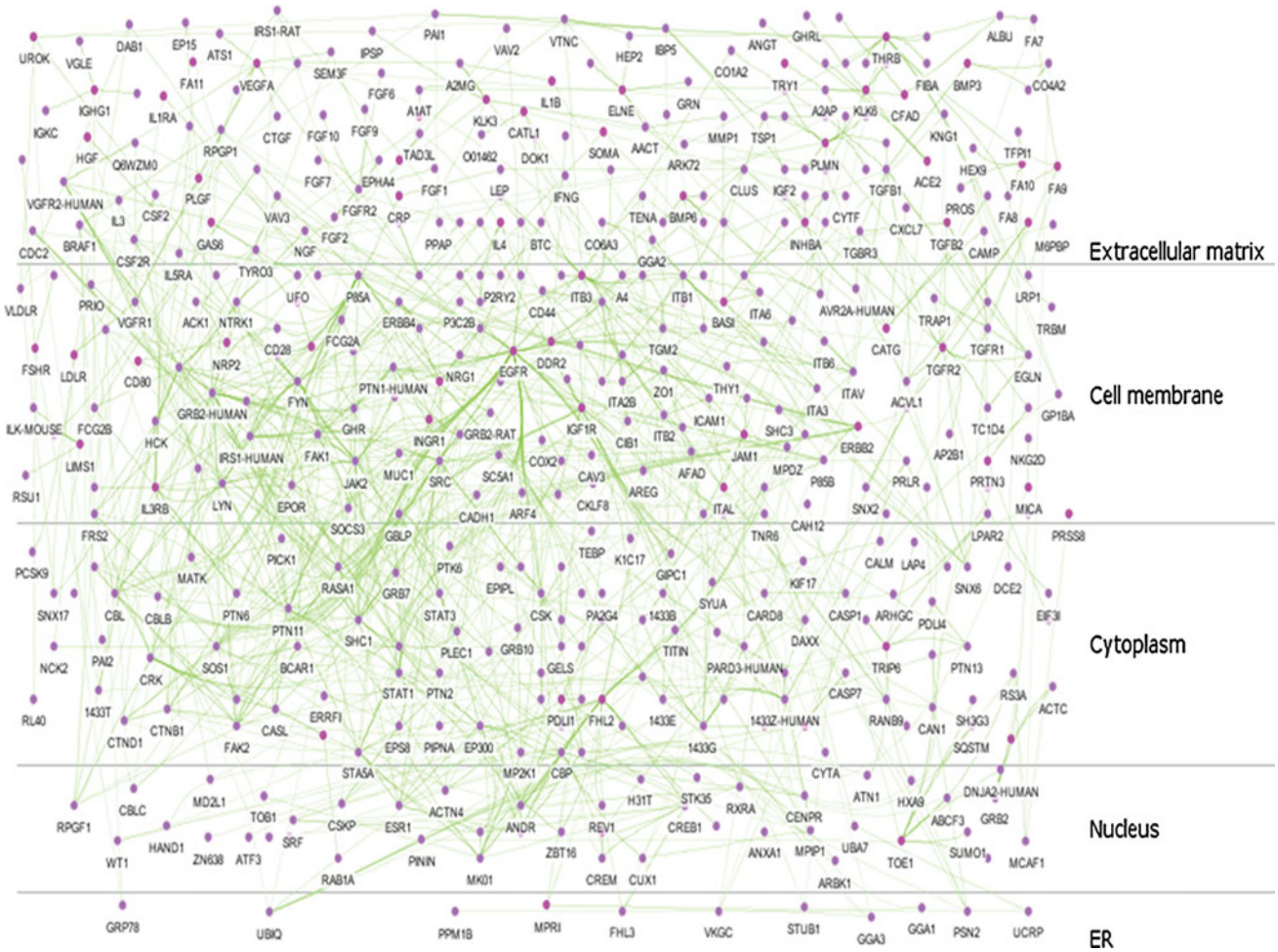

Fig. 4 Layered PPI network: 78 proteins and their interacting partners ($n = 445$) represented as *magenta* and *violet nodes* respectively (color figure online)

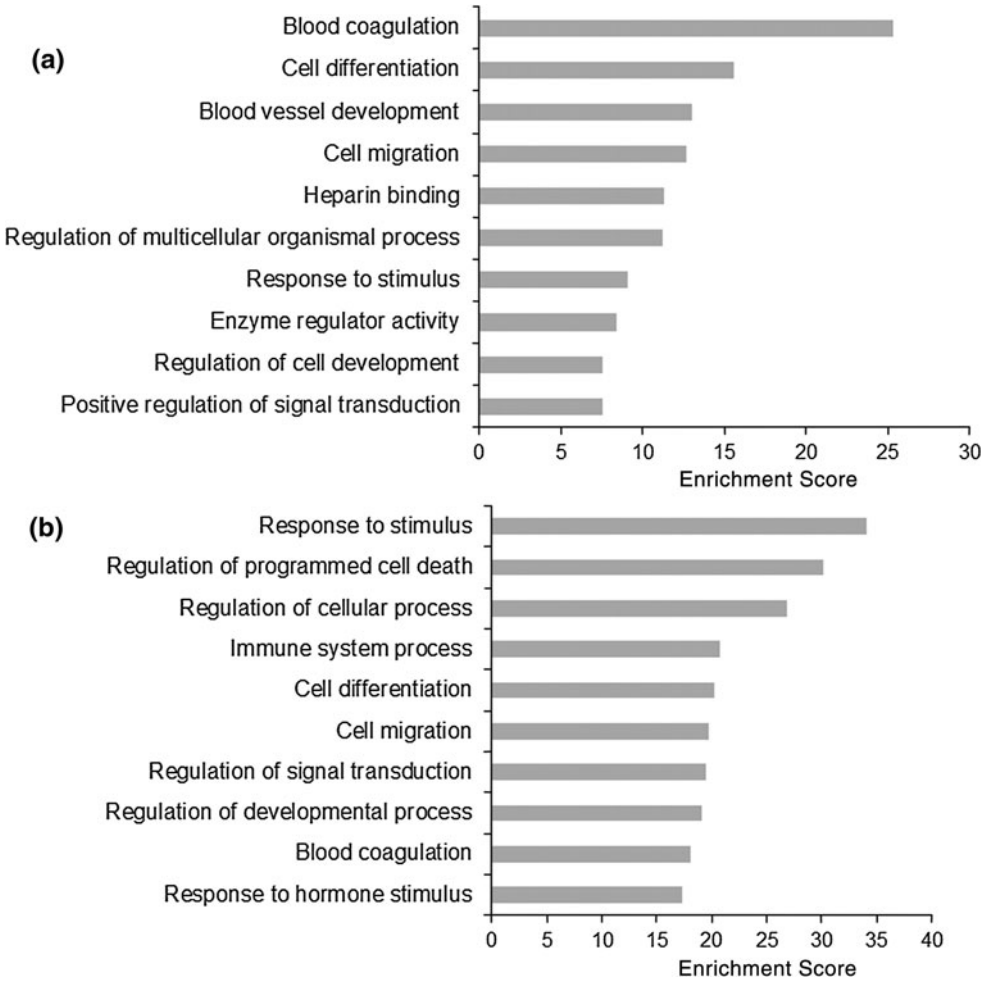

Fig. 5 Clustering of GO terms on PPI networks: significantly over represented top 10 clusters of biological process and molecular function of (*a*) 21 proteins and their interacting partners ($n = 174$) and (*b*) 78 proteins with their interacting partners ($n = 445$). Detailed information on statistical p value of significantly over represented the GO terms can be found in Table S5a and 5b, respectively

We also found that 330 cysteine residues in 130 proteins were free (in their reduced form). The pKa of these cysteine residues were determined using the PROPKA web server and H++ server (Anandakrishnan et al. 2012) Among the 330 cysteine residues, only 15 residues in 10 proteins were solvent accessible and had pKa lower than 8.3 as revealed by PROPKA server (Søndergaard et al. 2011) and also confirmed by H++ server (Table S3). This cut-off value was chosen since the pKa of cysteine is 8.3. Some of the cysteine residues that had low pKa's were found to be buried and hence were not considered as potential targets for Hcy.

Protein–protein interaction Network (PPI network)

Binding of small molecules like Hcy with a protein could potentially alter the dynamics of its protein–protein interaction. Hence, we tried to identify the interacting partners of the 21 proteins that have already been experimentally shown to bind with Hcy and the 78 proteins (that fitted the criteria mentioned above) using APID2NET (Hernandez-Toro et al. 2007) in Cytoscape (Shannon et al. 2003). PPI networks were modelled using three Cytoscape plugins viz Advance Merge networks, Network Analyzer (Assenov et al. 2008), and Cerebral (Barsky et al. 2007). The sub-cellular localization information was also imported from uniprot database and divided into four layers: extracellular, cell membrane, cytoplasm and nucleus. The first PPI network was built by considering only those interactions that have been validated in at least two experiments based on BioGrid, DIP, HPRD, IncAct and MINT databases (Hernandez-Toro et al. 2007). Among 21 proteins 6 did not have interactions that were validated by at least 2 experiments and hence for these proteins one experimental validation was also considered. Similar analysis was done for 12 of the 78 proteins identified in this study. In the 78 PPI network, 3 proteins have no information in APID server and 8 proteins were found as isolated nodes. The 21 protein PPI network was composed of 174 nodes and 274 interactions while the PPI network of 78 proteins had 445 nodes and 950 interactions (Figs. 3, 4). From the 21 PPI network it was shown that a majority of the homocysteinylated proteins (red nodes) and potential Hcy target proteins as identified in this study (magenta nodes) are found to be present in extracellular region and cell membrane.

Table 1 Significantly over represented top 10 pathways and disease association enrichment of 174 of 21 proteins PPI network

Category	Term	*p* value	Fold enrichment	Corrected *p* value (Benjamini)
KEGG_PATHWAY	hsa04610: complement and coagulation cascades	3.44E−17	13.904	3.23E−15
KEGG_PATHWAY	hsa04512: ECM-receptor interaction	3.28E−14	10.850	1.54E−12
KEGG_PATHWAY	hsa04510: focal adhesion	9.97E−13	5.966	3.12E−11
KEGG_PATHWAY	hsa05200: pathways in cancer	1.18E−09	3.948	2.78E−08
KEGG_PATHWAY	hsa04330: notch signalingz pathway	1.61E−09	12.248	3.02E−08
KEGG_PATHWAY	hsa04350: TGF-beta signalling pathway	9.74E−06	6.065	1.53E−04
KEGG_PATHWAY	hsa05215: prostate cancer	1.20E−05	5.929	1.61E−04
KEGG_PATHWAY	hsa05222: small cell lung cancer	4.91E−05	5.710	5.77E−04
KEGG_PATHWAY	hsa05220: chronic myeloid leukaemia	1.36E−04	5.756	0.0014164
KEGG_PATHWAY	hsa04520: adherens junction	1.64E−04	5.607	0.0015365
GENETIC_ASSOCIATION_DB_DISEASE	Thrombosis, deep vein	2.46E−14	17.841	2.33E−11
GENETIC_ASSOCIATION_DB_DISEASE	Thromboembolism, venous	4.97E−10	9.573	2.35E−07
GENETIC_ASSOCIATION_DB_DISEASE	Myocardial infarction	6.84E−10	7.603	2.16E−07
GENETIC_ASSOCIATION_DB_DISEASE	Cerebrovascular disease; sickle cell anaemia	2.08E−08	12.744	4.92E−06
GENETIC_ASSOCIATION_DB_DISEASE	Thrombophilia and vascular disease	3.17E−07	19.625	6.00E−05
GENETIC_ASSOCIATION_DB_DISEASE	Foetal loss, late	4.98E−07	14.018	7.85E−05
GENETIC_ASSOCIATION_DB_DISEASE	Thrombosis	3.00E−06	11.214	4.05E−04
GENETIC_ASSOCIATION_DB_DISEASE	Stroke	6.26E−06	6.1681	7.40E−04
GENETIC_ASSOCIATION_DB_DISEASE	Fibrinogen heart disease, ischaemic tissue plasminogen activator level	7.18E−06	28.037	7.54E−04
GENETIC_ASSOCIATION_DB_DISEASE	coronary artery disease	2.77E−05	4.7386	2.38E−03

Functional annotation and pathway analysis

GO terms

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used for the functional annotation of the first interacting partners of the 21 and 78 proteins. The significantly over represented GO terms identified extracellular region ($p = 1.00E{-}38$) as the most enriched cellular compartment (Table S4a) and these proteins are mainly involved in blood coagulation ($p = 2.83E{-}24$), cell differentiation ($p = 8.43E{-}18$), blood vessel development ($p = 3.31E{-}15$) and Cell migration ($p = 3.14E{-}13$) (Fig. 5a) (Table S5a).

The functional annotation clustering of 78 proteins PPI networks also identified extracellular regions proteins ($p = 1.37E{-}34$) (Table S4b) as the top cluster involved in various biological process such as response to stimulus ($p = 8.24E{-}32$), regulation of programme cell death ($p = 2.42E{-}30$), regulation of cellular process ($p = 1.60E{-}20$), cell differentiation ($p = 9.93E{-}20$), cell migration ($p = 8.81E{-}16$) and blood coagulation ($p = 1.32E{-}15$), etc. (Fig. 5b) (Table S5b).

Pathway enrichment and disease association

Following the functional annotation analysis, we mapped statistically overrepresented biological pathways and disease pathogenesis of these two PPI networks using 'KEGG pathway database' and GENETIC_ASSOCIATION_DB_DISEASE database option available in DAVID functional annotation tool. Among the top ten enriched pathways obtained using 21 and 78 proteins, complement and coagulation cascade, focal adhesion, ErbB signalling pathway, Pathways in cancer were found to be common (Tables 1, 2). Further, the interactome of the 21 and 78 proteins were found to be significantly associated with thrombosis, myocardial infarction, cerebrovascular disease, sickle cell anaemia, thromboembolism and stroke, respectively (Tables 1, 2).

Discussion

Molecular targeting of proteins by Hcy (S-homocysteinylation) leads to the formation of stable covalent disulfide bonds with cysteine residues. Homocysteinylation of

Table 2 Significantly over represented top 10 pathways and disease association enrichment of 445 of 78 proteins PPI network

Category	Term	*p* value	Fold enrichment	Corrected *p* value (Benjamini)
KEGG_PATHWAY	hsa04610: complement and coagulation cascades	3.38E−19	7.76	3.65E−17
KEGG_PATHWAY	hsa04510: focal adhesion	4.88E−19	4.35	2.64E−17
KEGG_PATHWAY	hsa04012: ErbB signalling pathway	6.74E−16	6.15	2.40E−14
KEGG_PATHWAY	hsa04630: Jak-STAT signalling pathway	1.53E−15	4.49	4.20E−14
KEGG_PATHWAY	hsa05200: pathways in cancer	6.17E−15	3.10	1.34E−13
KEGG_PATHWAY	hsa04650: natural killer cell-mediated cytotoxicity	3.28E−12	4.29	5.91E−11
KEGG_PATHWAY	hsa05220: chronic myeloid leukaemia	4.30E−11	5.47	6.64E−10
KEGG_PATHWAY	hsa05211: renal cell carcinoma	5.77E−10	5.35	7.79E−09
KEGG_PATHWAY	hsa04060: cytokine–cytokine receptor interaction	8.73E−09	2.72	1.05E−07
GENETIC_ASSOCIATION_DB_DISEASE	Thromboembolism, venous	1.65E−07	4.96	2.94E−04
GENETIC_ASSOCIATION_DB_DISEASE	Systemic lupus erythematosus	5.99E−07	3.09	5.35E−04
GENETIC_ASSOCIATION_DB_DISEASE	Inflammatory bowel disease	9.29E−07	4.43	5.53E−04
GENETIC_ASSOCIATION_DB_DISEASE	Myocardial infarction	1.24E−06	3.88	5.53E−04
GENETIC_ASSOCIATION_DB_DISEASE	Cerebrovascular disease; sickle cell anaemia	2.00E−06	6.36	7.14E−04
GENETIC_ASSOCIATION_DB_DISEASE	Thrombosis, deep vein	2.00E−06	6.36	7.14E−04
GENETIC_ASSOCIATION_DB_DISEASE	Arthritis	2.24E−06	3.93	6.67E−04
GENETIC_ASSOCIATION_DB_DISEASE	Celiac disease	7.89E−06	3.60	0.002
GENETIC_ASSOCIATION_DB_DISEASE	Stroke, ischaemic	1.44E−05	3.64	0.003
GENETIC_ASSOCIATION_DB_DISEASE	Stroke	1.61E−05	3.82	0.003

proteins has been reported to modulate functions of some proteins. For instance, binding of Hcy with fibronectin inhibits its ability to bind fibrin (Majors et al. 2002). Although molecular targeting hypothesis has been proposed as one of the mechanisms to explain the association of Hcy with a multitude of diseases, to date only 21 human proteins (Sundaramoorthy et al. 2008; Minagawa et al. 2010; Tang et al. 2011) have been experimentally proven to bind with Hcy. All these proteins have been identified using a hypothesis-based candidate protein approach. However, attempts to identify putative proteins that could bind Hcy are rare. We had earlier identified proteins that could bind Hcy based on high cysteine content (>5 %) of the protein. However, this limited the number of proteins identified that could potentially be homocysteinylated since proteins with cysteine content lower than 5 % could also bind to Hcy. In this study we identified proteins that have structural homology to the experimentally identified Hcy-binding proteins. These proteins have solvent accessible cysteine residues with low pKa (for free cysteine) or high DSE (for cysteine disulfides). Our analysis led to the identification of 78 proteins that we believe could be potential molecular targets for Hcy. The parameters that we chose for identifying proteins that could potentially bind to homocysteine, viz. solvent accessibility, dihedral strain energy of disulfide bonds or pKa of free cysteine residues, could be used as general parameters and we believe that using this approach one can successfully find if homocysteine (or for that matter other thiols) could potentially bind to any protein even if it is not structurally related.

Till date there are no studies identifying the interacting partners of proteins that bind or are predicted to bind Hcy. In this study we found that the 174 proteins directly interact with proteins that have been reported to bind Hcy while 445 proteins interact with the 78 proteins identified in this study. Interestingly, a larger proportion of these 78 proteins are extracellular and membrane proteins. Additionally, the interacting partners of the 21 homocysteinyted were also found predominantly enriched in extracellular region and involved in blood coagulation. The 445 interacting partners of the 78 potential Hcy target proteins (identified in this study) were also enriched in extracellular region. Apart from the coagulation pathway, we found that proteins involved in focal adhesion, ECM-receptor, ErbB signalling and cancer pathways having solvent accessible, high dihedral strain energy bonded cysteines or low pKa of cysteines could be potential Hcy target proteins as identified in our study. We also found that the alteration of these pathways are associated with various disease such as thrombosis, cerebrovascular, sickle cell anaemia, myocardial infarction, foetal loss and stroke as revealed by disease association enrichment analysis from DAVID. Most of these diseases have been shown to be associated with elevated levels of Hcy also (Eikelboom et al. 1999; Kirke et al. 1997; Perry et al. 1995; de Vries et al. 1997; Seshadri

et al. 2002; Meigs et al. 2001; Regland et al. 1995). Thus, we further postulate that binding of Hcy with any of these identified proteins (which we identified could be potential Hcy target protein) might disturb the protein–protein interaction and alteration of the pathways that are associated with various diseases. We believe that the identified targets would serve as a good starting point for validating proteins that undergo S-homocysteinylation modification.

Acknowledgments We thank all members of the lab for their valuable suggestions. Y.S. is thankful to CSIR-Human Resource Development Group for fellowship.

Conflict of interest The authors declare no conflict of interest.

References

Ahamed J, Versteeg HH, Kerver M, Chen VM, Mueller BM et al (2006) Disulfide isomerization switches tissue factor from coagulation to cell signaling. Proc Natl Acad Sci USA 103: 13932–13937

Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res 40:W537–W541

Andersson A, Lindgren A, Hultberg B (1995) Effect of thiol oxidation and thiol export from erythrocytes on determination of redox status of homocysteine and other thiols in plasma from healthy subjects and patients with cerebral infarction. Clin Chem 41:361–366

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24:282–284

Barsky A, Gardy JL, Hancock RE, Munzner T (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics 23:1040–1042

Chuang CC, Chen CY, Yang JM, Lyu PC, Hwang JK (2003) Relationship between protein structures and disulfide-bonding patterns. Proteins 53:1–5

da Huang W, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57

da Huang W, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13

de Vries JI, Dekker GA, Huijgens PC, Jakobs C, Blomberg BM et al (1997) Hyperhomocysteinaemia and protein S deficiency in complicated pregnancies. Br J Obstet Gynaecol 104:1248–1254

Eikelboom JW, Lonn E, Genest J Jr, Hankey G, Yusuf S (1999) Homocyst(e)ine and cardiovascular disease: a critical review of the epidemiologic evidence. Ann Intern Med 131:363–375

Gelly JC, Joseph AP, Srinivasan N, de Brevern AG (2011) iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res 39:W18–W23

Gilbert HF (1995) Thiol/disulfide exchange equilibria and disulfide bond stability. Methods Enzymol 251:8–28

Hernandez-Toro J, Prieto C, De las Rivas J (2007) APID2NET: unified interactome graphic analyzer. Bioinformatics 23:2495–2497

Jacobsen DW, Catanescu O, Dibello PM, Barbato JC (2005) Molecular targeting by homocysteine: a mechanism for vascular pathogenesis. Clin Chem Lab Med 43:1076–1083

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2537–2577

Kirke PN, Mills JL, Scott JM (1997) Homocysteine metabolism in pregnancies complicated by neural tube defects. Nutrition 13:994–995

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797

Majors AK, Sengupta S, Willard B, Kinter MT, Pyeritz RE et al (2002) Homocysteine binds to human plasma fibronectin and inhibits its interaction with fibrin. Arterioscler Thromb Vasc Biol 22:1354–1359

Mansoor MA, Svardal AM, Ueland PM (1992) Determination of the in vivo redox status of cysteine, cysteinylglycine, homocysteine, and glutathione in human plasma. Anal Biochem 200: 218–229

Meigs JB, Jacques PF, Selhub J, Singer DE, Nathan DM et al (2001) Fasting plasma homocysteine levels in the insulin resistance syndrome: the Framingham offspring study. Diabetes Care 24:1403–1410

Minagawa H, Watanabe A, Akatsu H, Adachi K, Ohtsuka C et al (2010) Homocysteine, another risk factor for Alzheimer disease, impairs apolipoprotein E3 function. J Biol Chem 285:38382–38388

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Perry IJ, Refsum H, Morris RW, Ebrahim SB, Ueland PM et al (1995) Prospective study of serum total homocysteine concentration and risk of stroke in middle-aged British men. Lancet 346: 1395–1398

Regland B, Johansson BV, Grenfeldt B, Hjelmgren LT, Medhus M (1995) Homocysteinemia is a common feature of schizophrenia. J Neural Transm Gen Sect 100:165–169

Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G (1998) Homology modeling, model and software evaluation: three related resources. Bioinformatics 14:523–528

Schmidt B, Ho L, Hogg PJ (2006) Allosteric disulfide bonds. Biochemistry 45:7429–7433

Sengupta S, Wehbe C, Majors AK, Ketterer ME, DiBello PM et al (2001a) Relative roles of albumin and ceruloplasmin in the formation of homocystine, homocysteine-cysteine-mixed disulfide, and cystine in circulation. J Biol Chem 276:46896–46904

Sengupta S, Chen H, Togawa T, DiBello PM, Majors AK et al (2001b) Albumin thiolate anion is an intermediate in the formation of albumin-S-S-homocysteine. J Biol Chem 276: 30111–30117

Seshadri S, Beiser A, Selhub J, Jacques PF, Rosenberg IH et al (2002) Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N Engl J Med 346:476–483

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. J Chem Theory Comput 7(7):2284–2295

Sundaramoorthy E, Maiti S, Brahmachari SK, Sengupta S (2008) Predicting protein homocysteinylation targets based on dihedral strain energy and pKa of cysteines. Proteins 71: 1475–1543

Tang YS, Khan RA, Zhang Y, Xiao S, Wang M et al (2011) Incrimination of heterogeneous nuclear ribonucleoprotein E1 (hnRNP-E1) as a candidate sensor of physiological folate deficiency. J Biol Chem 286:39100–39115

Thangudu RR, Manoharan M, Srinivasan N, Cadet F, Sowdhamini R, Offmann B (2008) Analysis on conservation of disulphide bonds

and their structural features in homologous protein domain families. BMC Struct Biol 8:55

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta SJ, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784

Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC et al (2006) The universal protein resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187–D191

Zemla A (2003) LGA:a method for finding 3D similarities in protein structures. Nucleic Acids Res 31:3370–3374