Bulgarian Chemical Communications, Special Issue E, (pp. 50 – 53) 2017

In silico investigation of single-chain variable fragment (scFv) antibody, structurally similar to native C1q globular heads

M.A. Rangelov[1*], N.H. Todorova[2], I.G. Tsacheva[3*]

[1] *Institute of Organic Chemistry with Centre of Phytochemistry (IOCCP), Bulgarian Academy of Sciences, Acad. G. Bontchev str. bl. 9, 1113 Sofia, Bulgaria*

[2] *Institute of Biodiversity and Ecosystem Research (IBER), Bulgarian Academy of Sciences, Yurii Gagrin Str.2, 1113 Sofia, Bulgaria*

[3] *Biology Faculty, Sofia University "St. Kliment Ohridski", 8 Dragan Tsankov Blv., 1164 Sofia, Bulgaria*

Received September 29, 2016; Revised February 04, 2017

Homology model of a selected single-chain variable fragment (scFv) inhibitory antibody A1 is generated using Generalized Born/Volume Integral (GB/VI) methodology and MMFF94 force field. Comparative analysis of scFvA1 homology model and ghC1q chains is further used to find conformational and electrostatic similar regions, crucial for the biological binding regions in C1q globular heads.

Key words: homology modeling, anti-C1q scFv

INTRODUCTION

Phage-displayed single-chain variable fragment (scFv) antibodies obtain serious advantages that make them preferable tool for application in research, laboratory diagnostics and medicine. Recent progress in antibody engineering together with the microbial expression systems are milestones for design and production of antibodies for numerous applications [1]. Variety of selection and screening strategies could be used to simultaneously derive many high-affinity scFv with different relevance [2]. Importance of their diverse collection relies on high specificity and selectivity of antibodies, while offering distinct biological profiles in tissue distributions of unique molecules [3].

C1q is the recognizing first subcomponent of the classical complement pathway, which undergo a conformation transition by its activation, exposing neo-epitopes. Naive phage library is used to create a single-chain variable fragment (scFv) antibody, that mimics the globular head regions of native C1q (ghC1q). It is further used in in-silico modelling to characterise crucial for biological binding regions in C1q globular heads.

EXPERIMENTAL

scFv A1 construction and sequencing

'Griffin' library, provided by Greg Winter, MRC, Cambridge, UK, is used to create large, nonimmunized fully human scFv repertoires [4], with high specificity of binding to human self-antigens [5]. This kind of library construction of naturally rearranged V genes in a phagemid vector ensures natural diversity in length of the VH CDR3 and a higher number of functional scFv [2]. ScFv A1, showing functional mimicry with only to the globular head region of C1q, is obtained after several panning rounds. pHEN2 region, containing genes for scFv A1 is amplifyed by using in-situ polymerase chain reaction (PCR) of the Escherichia coli clone and primers LMB3 (5`-CAGGAAACAGCTATGAC-3`) and fd-SEQ1 (5`-GAATTTTCTGTATGAGG-3`), using amplification programm of Liners *et al.* 2005 [6]. The product is purfied and further send for iboth sides sequencing (Macrogen Europe). Derived sequence is deposited in GenBank under accession number KX981596 .

Homology Search

Searches for protein structures that are homologous to a query sequence are performed on a database of protein structures and sequences that have been clustered into families. A scan is performed to create an initial list of candidates using a generalized version of the FASTA methodology [7]. Expectation value (E-Value) is determined for each sequence and low E-Values, that correspond to good scores were used as evaluation criteria.

Homology Model of scFv A1

We performed alignment of the target sequence to the selected parts of best scored selected proteins according to the modified version of Needleman

* To whom all correspondence should be sent:
E-mail: marangelov@gmail.com

and Wunsch algorithm [8] with sequence alignment imposed on amino acid BLOSUM 62 substitution matrix. Homology Model is carried out by Boltzmann-weighted randomized modeling procedure using coulomb and generalized Born implicit solvent [9] interaction energies. Best models were further refined by Merck Molecular Forcefield MMFF94 [10] for all energy minimization.

Comparative analysis of scFvA1 homology model and ghC1q chains

Sequence comparison of crystallized globular heads domains of human C1q, according to the 1PK6 PDB and H-chain of generated homology model of scFv A1 is performed, using secondary structure information of those sequences having associated atomic coordinates. The target sequence is sequence-to-group aligned to each one of ghC1q chains. Manual alignment is used for refining CDR frames of scFv A1 H-chain and loops in A-, B- and C- ghC1q.

RESULTS AND DISCUSSION

Sequence and structure similarity analysis

Analysis of the CDR3 regions of scFv clones from the library, both VH and VL, shows considerable variability in their length, which is similar to the distribution of the length of CDR3 in unselected library.

ScFv A1 nucleotide sequence is translated to amino acids (Fig. 1), compared to deposited germ-line V genes and most likely CDR regions in both VL and VH were predicted in VBASE2 [11]. Heavy chain V segment is of the IGHV4 family, while the one of the light chain used IGKV1 subgroup. No J-segment is discovered in the VL, but the junction in VH participated as expected in the CDR3 formation

The scFv A1 model

Sequence alignment is performed using MOE [12] software with database deposited scFv antibodies by BLOSUM 62 algorithm as substitution scoring matrix, with algorithm penalty for gap opening -12 and gap extension -2 and evaluated with E-value accepted $1.e^{-12}$. The first five hits (PDB ID: 1YNL, 1A5F, 1ETZ, 1DEE, 1AD9) showing statistically significant similarity, the most similar one having $E \sim 2.e^{-42}$, were chosen for further analysis (Fig. 2).

Further multiple sequence alignment determines the correspondences between residues of selected related protein chains employing *sequence-derived* and *structure-based* information. BLOSUM 62 substitution matrix is chosen as based on observed alignments and immunoglobulin superfamily having many specific members.

>Amino acid sequence of H-chain scFv A1

RCSSRSRGAGLLKPSETLSLTCAVYGGSFSGYYWSWIRQPPGKGLEWIGEINHSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARSHSAAWGQGTLVTVS

>Amino acid sequence of L-chain scFv A1

DIQLTQSPSFLSASVGDRVTITCRASQGISSYLAWYQQKPGKAPKLLIYAASTLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYHCQQLNSY

CDR analysis of H-chain scFv A1

GGSFSGYY....___INHSGST...___ARSHSAA

----CDR1---> <--CDR2--> <---CDR3----

CDR analysis of L-chain scFv A1

QGISSY.....___AAS.....___QQLNSY

----CDR1---> <--CDR2--> <---CDR3----

Fig. 1. ScFv A1 H- and L-chain amino acid sequence, derived from nucleotide sequence, using standard genetic code in VBASE2, with predicted CDR 1, 2 and 3 in H- and L-chains.

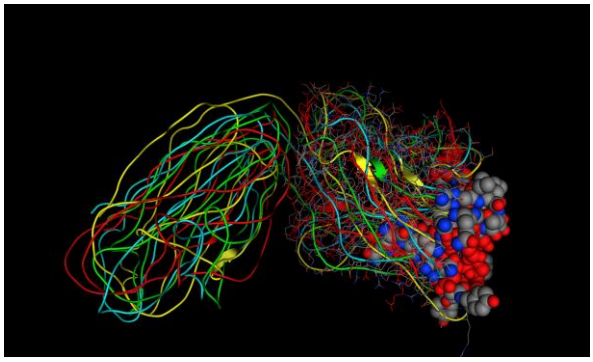

Fig. 2. Comparison of scFv VL and VH structure to high scored similar PDBs of crystallized antibodies, shown in different colors. CDR1, 2 and 3 of VH is visualized as CPK.

Homology model of scFv A1 is generated using Generalized Born/Volume Integral (GB/VI) methodology and MMFF94 force field (Fig. 3).

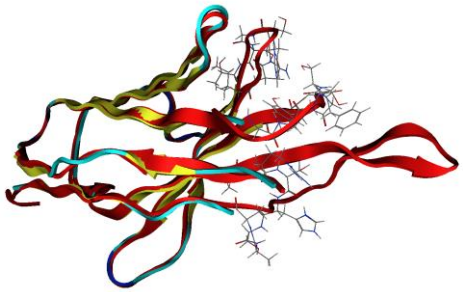

Fig. 3. Representation of the known 3D structure of crystallized antibodies in red and the generated model backbone of scFv A1 in color coding (yellow beta-structure, blue for loops).

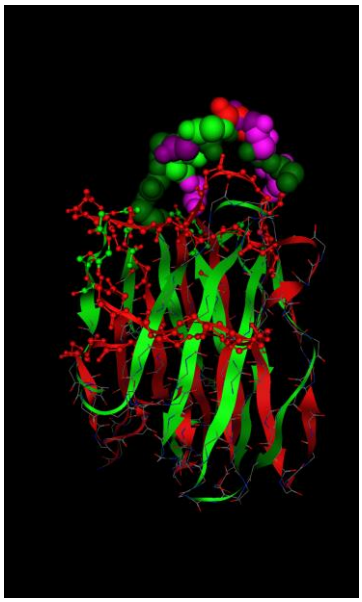

Fig. 4. CDR1 H-chain scFv A1 according to the model: GGSFSGYY dark color and A-chain of C1q: TTNKGLF; light color.

ScFv A1 H-chain and ghC1q comparison and loop similarity analysis

ScFv A1 CDR1, 2 and 3 were compared loop by loop with C1q separate chains loop. Similarity in using AA coding is: polar + and – separately; polar uncharged and nonpolar. Backbone is analyzed only as amino acid side chain (R) polarity due to the mobility of R in solution, and not as a chain position in the loop. The following color coding is formed according to amino acid R-chain properties analysis (Table 1):

Table 1. Good correlation of R polarity between C1q globular head chains and CDR loops of H-chain scFv A1 is achieved for B-chain and CDR2, A-chain and CDR1 and C-chain and CDR3.

CDR1	H		G	G	S	F	S	G	Y	Y	
Ac1q		55-60	L	S	Q	W	E	I	C	L	
Bc1q		82-91	D	Y	A	Y	N	T	F	Q	V
Cc1q		79	H	T	S	K	T	N	Q	V	
CDR2	H		I	N	H	S	G	S	T		
Ac1q		121-128	D	A	E	S	G	Q	Y	I	
Bc1q		116-123	G	N	E	G	A	N	S	I	
Cc1q		55-63	A	S	H	T	A	N	L	C	
CDR3	H		A	R	S	H	S	A	A		
Ac1q		81-88	T	T	N	K	G	L	F	Q	
Bc1q		36-43	A	S	S	R	G	N	L	C	
Cc1q		112-119	V	G	I	Q	G	S	D	S	

R,H,K -> positively charged in darker gray; D,E -> negatively charged in dark gray; S,T,N,Q,P,C -> polar uncharged (H-bonds) in gray; G,A,V,I,L,M,F,Y,W -> hydrophobic in light gray.

CONCLUSIONS

Based on the nucleotide sequence of an inhibitory scFv A1, a structural 3D model of resulting polypeptide chain is generated, as homology model according to alignment on the deposited crystallized scFv PDBs. The resulting structure is used to find conformational (Fig. 4) and electrostatic similar regions between homology structure and native human C1q globular heads. This similarity helps to clarify crucial recognising regions from C1q molecule. The presented model will be further helpful giving insights into new scFv binding properties that represent a step toward understanding the recognition mechanisms of C1q and their biological targets.

REFERENCES

1. Z.A. Ahmad, S.K. Yeap, A.M. Ali, W.Y. Ho, N.B.M. Alitheen, M. Hamid, *Clin Dev Immunol.*, **2012**, 1 (2012).
2. S. Lennard, Standard Protocols for the Construction of scFv Libraries, Vol. 178 Methods in Molecular Biology Series, Humana Press, Inc., Tutowa NJ, 2002.
3. A.L. Nelson, *Mabs*, **2**, 77 (2010).
4. A.S. Zlatarova, I. Tsacheva, M.S. Kojouharova, *Biotechnol. & Biotechnol. Eq.*, **19**, 2 (2005).

5. A.D. Griffiths, M. Malmqvist, J.D. Marks, J.M. Bye, M.J. Embleton, J. McCafferty, M. Baier, K.P. Holliger, B.D. Gorick, N.C. Hughes-Jones, H.R. Hoogenboom, G. Winter, *The EMBO Journal*, **12**, 725 (1993).
6. F. Liners, W. Helbert, P. Van Cutsem, *Glycobiology*, **15**, 849 (2005).
7. W. R. Pearson, *Meth. Enz.*, R. F. Doolittle, ed. (San Diego: Academic Press) **266,** 227 (1996).
8. S.B. Needleman, C.D. Wunsch, *Journal of Molecular Biology*, **48**, 443 (1970).
9. P. Labute, *J. Comp. Chem.*, **29**, 1693 (2008).
10. T.A. Halgren, *J. Comput. Chem.*, **17**, 490 (1996).
11. I. Retter, H.H. Althaus, R. Münch, W. Müller, *Nucleic Acids Res.*, **33** (Database issue), D671 (2005).
12. Molecular Operating Environment (MOE), 2011.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2011.

In silico ИЗСЛЕДВАНЕ НА ЕДНОВЕРИЖНИ АНТИТЕЛА (scFv), СТРУКТУРНО ПОДОБНИ НА НАТИВНИТЕ ГЛОБУЛАРНИ ФРАГМЕНТИ НА C1q

М.А. Рангелов[1*], Н.Х. Тодорова[2], И.Г. Цачева[3*]

[1] Институт по органична химия с Център по фитохимия (ИОХЦФ), Българска академия на науките, ул. "Акад. Г. Бончев" бл. 9, 1113 София, България
[2] Институт по биоразнообразие и екосистемни изследвания (ИБЕИ), Българска Академия на Науките, ул. "Юрий Гагарин" 2, 1113 София, България
[3] Биологически факултет, Софийски университет "Св. Климент Охридски", бул. "Драган Цанков" 8, 1164 София, България

Постъпила на 29 септември 2016 г.; Коригирана на 4 февруари, 2017 г.

(Резюме)

Генериран е хомоложен модел на избрано едноверижно инхибиращо антитяло (single-chain variable fragment (scFv) A1 посредством алгоритъма Generalized Born/Volume Integral (GB/VI) и MMFF94 силово поле. Анализиран е хомоложния модел на scFvA1 спрямо глобуларните фрагменти на C1q (ghC1q) резултатите от анализа са използвани за откриване на конформационни и електростатични региони на подобие, критични за биологичните функции на свързване в глобуларните региони на C1q.